(S)electing subsets

CPS 290.4/590.4: Crowdsourcing Societal Tradeoffs

April 8, 2015

Saša Pekeč
Decision Sciences
The Fuqua School of Business
Duke University

pekec@duke.edu
http://people.duke.edu/~pekec

Representation

- limited resources:
  - single representative / choice
  - multiple representatives / set of choices

- “optimal” resource allocation:
  - objective(s):
    - winner-takes-all
    - maximizing total welfare/value
    - minimizing regret/loss/cost
    - fairness?

Representation

- parliaments, assemblies
- committees, company boards, project teams
- product assortment, product bundles

Stylized (simplistic) setting:

- m candidates: $a,b,c,\dots$
- n single-minded choices $n = n_a + n_b + n_c + \dots$

  (w.l.o.g. assume $n_a \geq n_b \geq n_c \geq \dots$)
- k slots to be assigned to m candidates

Proportional Representation

Total of *n* choices for *k* slots: *n/k* choices per slot

Quota:

Assign candidate *j* either

$LQ = INT(n_j/(n/k))$ or $UQ = 1 + INT(n_j/(n/k))$

- **The method of largest reminders**
  (a.k.a., Hamilton's method, Vinton's method)

  1. assign each candidate *j*: $LQ = INT(n_j/(n/k))$ slots

  2. assign remaining $n - \sum INT(n_j/(n/k))$ slots to candidates with largest reminder values

  $$n_j/(n/k) - INT(n_j/(n/k))$$

US Constitution Article I, Section 2

- …the House Representatives be *apportioned* among the several States according to their respective populations;
- …the number of Representatives shall not exceed one for every 30,000 persons;
- …each State shall have at least one Representative;
- …the reapportionment shall occur once every 10 years as a result of the decennial census.

Apportionment of the seats in the US House of Representatives

Fairness of the largest reminders approach?

- **The Alabama Paradox**
- **The New State Paradox**
- **The Population Paradox**

The Alabama Paradox

Apportionment based on 1880 Census:

With 299 seats: Alabama gets 8 seats (quota of 7.646)

With 300 seats: Alabama gets 7 seats (quota of 7.671)

A candidate (state) can lose a slot (seat) if the number of slots increases.

- other reminders become larger
- could construct smaller examples (with three candidates)

The New State Paradox

Oklahoma became a state in 1907. The house had 386 seats. Given Oklahoma population adding to the US population total at that time, Oklahoma should get 5 seats and the House should increase to 386+5=391 seats.

When the method of largest reminders was reapplied, Maine moved from 3 to 4 seats, while New York moved from 38 to 37 seats.

Adding a new candidate and additional choices supporting the candidate (without changing support for others), slot assignments of existing candidates can change.

The Population Paradox

According to 1900 Census, Virginia had a significantly higher (absolute and percentage) population growth than Maine.

	1900	1901 (est.)
Virginia	1,854,184	1,873,951 (1.07% increase)
Maine	694,466	699,114 (0.67% increase)
Virginia	**10** (9.599 quota)	**9** (9.509 quota)
Maine	**3** (3.595 quota)	**4** (3.548 quota)

Candidate whose support increases at a faster (absolute and percentage) rate could lose slots to a candidate whose support increases at a slower rate.

An Impossibility Theorem

Theorem (Balinski & Young 1982)

There exists no method that allocates k slots to $m>2$ candidates and

(1) satisfies Quota,

(2) avoids the Alabama Paradox,

(3) avoids the New State Paradox,

(4) avoids the Population Paradox.

- In fact the theorem holds with (1) and (4) only (Also, with (1) and (3) only.)
- If one can choose $k' \geq k$, then a method can be constructed
- Plenty of methods satisfying (2),(3) and (4).

Divisor methods

- Instead of specifying *n/k* rate per slot, choose rate *d*

1. Pick a rounding rule $r: \boldsymbol{N} \rightarrow \boldsymbol{R}$,

if $x < r(INT(x))$ then round down to $INT(x)$

if $x \geq r(INT(x))$ then round up to $1+INT(x)$

2. Pick divisor *d*

3. Assign slots to each candidate *j* according to $r(INT(n_j/d))$

- Pick divisor *d* so that exactly *k* slots are assigned.

(not unique)

- Divisor methods avoid all three paradoxes by construction
- For any divisor method there is an instance that violates quota

Webster's method

(a.k.a. Sainte-Laguë, major fractions)

1. Use a standard rounding rule r: $\boldsymbol{N} \rightarrow \boldsymbol{R}$,

if $x < INT(x)+0.5$ then round down to $INT(x)$

if $x \geq INT(x)+0.5$ then round up to $1+INT(x)$

2. Pick divisor d

3. Assign slots to each candidate j according to $r(INT(n_j/d))$

Example: k=4; $n_a= 61$, $n_b= 14$, $n_c= 13$, $n_a= 12$.

Use $d=26$ (but could also use $d=25$)

$a \rightarrow 61/26 = 2.35 \rightarrow 2$ slots,

$b \rightarrow 14/26 = 0.54 \rightarrow 1$ slot,

$c \rightarrow 13/26 = 0.50 \rightarrow 1$ slot,

$d \rightarrow 12/26 = 0.46 \rightarrow 0$ slots.

Jefferson's method

(a.k.a. d'Hondt, greatest divisors)

1. Always round down r: $\boldsymbol{N} \rightarrow \boldsymbol{R}$,

if $x < INT(x)+1$ then round down to $INT(x)$

if $x \geq INT(x)+1$ then round up to $1+INT(x)$

2. Pick divisor d

3. Assign slots to each candidate j according to $r(INT(n_j/d))$

Example: k=4; $n_a = 61$, $n_b = 14$, $n_c = 13$, $n_a = 12$.

Use $d=15$

$a \rightarrow 61/15 = 4.03 \rightarrow 4$ slots,

$b \rightarrow 14/15 = 0.93 \rightarrow 0$ slots,

$c \rightarrow 13/15 = 0.86 \rightarrow 0$ slots,

$d \rightarrow 12/15 = 0.8 \rightarrow 0$ slots.

Huntington-Hill's method

(a.k.a. equal proportions)

1. Geomean-based rounding rule: r: $\boldsymbol{N} \rightarrow \boldsymbol{R}$,

if $x < SQRT((INT(x)*(INT(x)+1))$ then round down to $INT(x)$

if $x \geq SQRT((INT(x)*(INT(x)+1))$ then round up to $1+INT(x)$

2. Pick divisor d

3. Assign slots to each candidate j according to $r(INT(n_j/d))$

Example: k=4; $n_a = 61$, $n_b = 14$, $n_c = 13$, $n_a = 12$.

Use $d=50$

$a \rightarrow 61/50 = 1.22\ (r=1.41) \rightarrow 1$ slot,

$b \rightarrow 14/50 = 0.28\ (r=0) \rightarrow 1$ slot,

$c \rightarrow 13/50 = 0.26\ (r=0) \rightarrow 1$ slot,

$d \rightarrow 12/50 = 0.24\ (r=0) \rightarrow 1$ slot.

Turbulent US history

- All four methods have been used for apportionment of US House seats at different times
- Cause of the very first presidential veto (1791)
- Cause of the incorrect result of the presidential election in 1876 (Hayes vs. Tilden, due to incorrect apportionment four years earlier)
- Cause of a direct violation of the constitution in 1921
- Huntington-Hill used since 1941, survived the constitutional challenge (upheld by the Supreme Court in 1992)

Divisor methods

- Jefferson's (d'Hondt) favors large n_j
  - desirable property for seat assignment in parliamentary elections? (majority building), used in many countries
- Webster (Sainte-Laguë) is claimed to introduce least bias, used in many countries (e.g., Germany)
- Huntington-Hill:
  -guarantees a slot to each candidate
  -minimizes pairwise discrepancies in the number of votes per slot among candidates
- Intense debate (and theorems) on
  Webster vs. Huntington-Hill
  (arithmetic vs. geometric mean)

Other Proportional Representation Methods

- Winner-takes-all: assign all k slots to candidate a
- Generate n_a, n_b, n_c, ... by other methods: scoring, weighted voting, approval, etc.
- Minimize misrepresentation (Monroe 1995, Chamberlain and Courant 1983):
  - Assign voters to each of k slots.
  - For each voter calculate an individual degree of misrepresentation given candidate allocated that slot
  - Aggregate total misrepresentation for a slot (sum of individual misrepresentations, weighted by #voters assigned to a slot)
- Can be described as a huge IP (Potthoff and Brams, 1998), NP-complete (Proccacia et al, 2008)

Subset (S)election

- Proportional representation:

  k slots (seats) > m candidates (parties, states)

- **What if $k < m$?**

  Standard (s)election/choice arguments/methods impossibility results apply.

  Can think of all possible 2^m subsets (or just allowed ones, e.g., all subsets of prespecified size k) as available alternatives and elicit preferences over those.

  $\Rightarrow$ Exponential blow-up

  However, new combinatorial issues emerge.

Subset (S)election

$\Rightarrow$ choosing a subset **S**
from the set of ***m*** available alternatives

$\Rightarrow$ choosing a feasible (admissible) subset **S**

- **social choice**
- **voting**
- **multi-criteria decision-making**
- **consumer choice (?)**

Social Choice

There are ***n*** individuals, each having preferences over ***m*** alternatives.

- **How to aggregate preferences into a "consensus" preference structure?**

- **Arrow's Impossibility Theorem**

  - Independence of Irrelevant Alternatives (IIA)

  - Inherent multidimensionality (single-peaked prefs.)

Choosing a subset?

Voting

There are ***n*** voters, each having preferences over ***m*** candidates.

- **How to aggregate preferences and determine the winner?**

- **Gibbard-Satterhwaite Theorem**

  - IIA implying Arrow's result

  or

  - manipulability

Choosing a subset?

Multi-Criteria Decision-Making (MCDM)

There are **n** criteria, each defining a preference structure over **m** alternatives

- **How to aggregate preferences and determine the "consensus" preference structure over alternatives?**

Choosing a subset?

Consumer Choice?

- **Behavioral aspects dominate**

(Normative approach: multicriteria decision-making)

BUT…

Automated consumer choice suggestions

- pagerank in search results, ad placement, etc.
- suggested product (e.g., credit cards, computers)

….

Choosing a subset?

Consumer Choice?

- **buying a product bundle**

  - related products (e.g., media system components)

  - subset of extra options (cars)

  - features of a highly customized commodity-like products (laptops, smartphones, software,…)

- **multiple criteria** (functionality, looks, safety, price, …)

Choosing a single alternative

Information requirement on voters' preferences

- SWF $\Rightarrow$ rankings
- plurality $\Rightarrow$ top choice
- prediction, betting, scoring rules $\Rightarrow$ constrained cardinal utility (IIA???)
- **approval voting $\Rightarrow$ subset choice**

CHOSING A SINGLE ALTERNATIVE

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1	2	8	3	6	1	7	8	5	5
2	1	7	8	1	3	1	7	1	6
3	7	1	5	4	2	8	5	2	2
4	5	4	1	8	8	3	4	8	7
5	4	3	6	3	7	4	3	3	4
6	6	5	7	2	6	5	1	4	8
7	8	2	2	7	5	6	2	6	1
8	3	6	4	5	4	2	6	7	3

CHOSING A SINGLE ALTERNATIVE

APPROVAL

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1	2	8	3	6	1	7	8	5	5
2	1	7	8	1	3	1	7	1	6
3	7	1	5	4	2	8	5	2	2
4	5	4	1	8	8	3	4	8	7
5	4	3	6	3	7	4	3	3	4
6	6	5	7	2	6	5	1	4	8
7	8	2	2	7	5	6	2	6	1
8	3	6	4	5	4	2	6	7	3

CHOSING A SINGLE ALTERNATIVE

APPROVAL

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1		✓				✓	✓		
2		✓	✓				✓		
3	✓					✓			
4				✓	✓			✓	✓
5					✓				
6		✓							✓
7	✓			✓					
8		✓						✓	

CHOSING A SINGLE ALTERNATIVE

APPROVAL

	V1	V2	V3	V4	V5	V6	V7	V8	V9	
1	**0**	**1**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**3**
2	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**3**
3	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**2**
4	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**1**	**1**	**4**
5	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**
6	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**2**
7	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**2**
8	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**2**

APPROVAL VOTE PROFILE

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1		✓				✓	✓		
2		✓	✓				✓		
3	✓					✓			
4				✓	✓			✓	✓
5					✓				
6		✓							✓
7	✓			✓					
8		✓						✓	

37, 1268, 2 , 47 , 45 , 13 , 12 , 48 , 46

APPROVAL VOTE PROFILE

	V1	V2	V3	V4	V5	V6	V7	V8	V9	
V=(	**37,**	**1268,**	**2**	**, 47 ,**	**45 ,**	**13 ,**	**12 ,**	**48 ,**	**46**	**)**

CHOSING A SUBSET

Information requirement on voters' preferences

- SWF $\Rightarrow$ rankings on all 2^m subsets
- plurality $\Rightarrow$ top choice among all 2^m subsets
- scoring rules $\Rightarrow$ constrained card. utility on all 2^m subsets
- approval voting $\Rightarrow$ subset choice on all 2^m subsets

CHOSING A SUBSET

- using “consensus” ranking of alternatives

  but

  for all voters: 1>2>3 or 2>1>3

  AND

  13>23>12 or 23>13>12

- divide and conquer:

  break into several separate singleton choices

- proportional representation

IGNORING INTERDEPENDANCIES
(substitutability and complementarity)

CHOSING A SUBSET

- Barbera et al. (ECA91): impossibility

A manageable scheme that accounts for interdependencies?

Proposal: Approval Voting with modified subset count.

Threshold Approach:

- define *t(S)* for every feasible S

- $AC_t(S)$= # of voters *i* such that $|V_i \cap S| \geq t(S)$

AV THRESHOLD APPROACH

Define t(S) for every feasible S

$AC_t(S)$= # of voters i such that $V_iS = |Vi \cap S| \geq t(S)$

Threshold functions (TF):

- t(S)=1 (minimal representation)
- t(S) = |S|/2 (majority)
- t(S) = (S|+1)/2 (strict majority)
- t(S) = |S| (unanimity)

….

COMPLEXITY of AVCT

- If X= the set of all feasible subsets, is part of the input then computing AVCT winner is polynomial in mn+|X|

Theorem.

If X is predetermined (not part of the input), then computing AVCT winner is NP-complete at best.

Proof: choosing a k-set, $t \equiv 1$. Suppose $|V_i|=2$ for all i.

Note: alternatives ~ vertices of a graph

V_i ~ edges of a graph

k-set approved by all voters ~ vertex cover of size k

Vertex Cover is a fundamental NP-complete problem.

COMPLEXITY cont'd

- not as problematic as it seems.

Theorem. (Garey-Johnson)

If X is predetermined (not part of the input), then computing

$$\mathbf{max_{S\ in X}\ sum_{i\ in S}\ score(i)}$$

is NP-complete.

LARGER IS NOT BETTER

Example: m=8, n=12, strict majority TF: t(S)=(|S|+1)/2

V= (123,15,1578,16,278,23,24,34,347,46,567,568)

- 1-set (AC):

1,2,3,4,5,6,7 all approved by <u>4 voters</u> (8 is approved by 3 voters)

- 2-set:

15,23,34,56,57,58,78 all approved by <u>2 voters</u>

- 3-set: 234 approved by <u>5 voters</u>
- 4-set: 5678 approved by <u>3 voters</u>
- 5-set: 15678 approved by <u>4 voters</u>

TOP INDIVIDUAL NOT IN A TOP TEAM

Example: m=5, n=6, majority TF: t(S)=|S|/2

V= (123,124,135,145,25,34)

- Top individual:

1 approved by <u>4 voters</u> (all other alternatives approved by 3 voters)

- Top team

2345 is the only team approved by all <u>5 voters</u>

- could generalize examples for almost any TF

- could generalize to top k individuals

THRESHOLD SENSITIVITY

Theorem

For any K>1, there exist n,m and a corresponding V such that AVCT winner S_k (where X is the set of all K-sets), k=1,…,K are mutually disjoint.

ANY GOOD PROPERTIES?

P1. Nullity.

If every vote is the empty set, any choice is good.

P2. Anonymity.

If U is a permutation of V, the choices for U and V are identical.

P3. Partition Consistency.

If S is chosen in two voter disjoint elections, then S would be chosen in the joint election.)

P4. Partition Inclusivity.

If no S is chosen by a single voter and in an election of the remaining n-1 voters, then any choice would also be chosen in an election w/o one of the voters.

SINGLE VOTER PROPERTIES

$\tau(S) = \min\{AS: S \text{ is a choice for } A\}$

P5. For every choice S, there exists votes A and B such that A is a choice for S but not for B.

P6. Let S be a choice for vote A that does not choose everyone. If BS>AS then S is a choice for B

P7. For every S, there is an A such that $AS = \tau(S) - 1$

P8. Suppose vote B chooses every committee. For all A_1, A_2 and for all choices S, T: If $A_1S = \tau(S)$, $A_2T = \tau(T)$, then $BS > A_1S$ implies $BT > A_2T$

Characterization Theorem

Theorem (Fishburn and P.)

If P1-8 hold, then the subset choice function is the AVCT.

AV THRESHOLD APPROACH

- low informational burden

- simplicity

- takes into account subset preferences

Results:

- properties of TFs, axiomatic characterization
- complexity
- robustness properties: theorems show what is possible and not what is probable

Need:

-Comparison with other methods, data validation
- strategic considerations

(S)electing subsets: Summary

n voters, *m* candidates, *k* slots

Proportional representation (m<k)

-quota, paradoxes, impossibility theorems

-multiple methods used in practice

Subset selection (m>k)

-standard voting and social choice results extend

-new combinatorial issues

-approval voting is a natural method (since it is subset-based)

-threshold approach reduces informational burden

-open field for other methods