Probabilistic studies of election methods

Svante Janson

KTH, Stockholm, 30 May 2011

Notation

We have m parties with v_i votes for party i; $V := \sum_{i=1}^{m} v_i$ is the total number of votes and $p_i := v_i/V$ the proportions of votes for party i.

The house size is n and party i gets s_i seats; thus

$$\sum_{i=1}^{m} s_i = n.$$

Seat excess and bias

Strict proportionality would give

$$q_i := \frac{v_i}{V} n = p_i n \tag{1}$$

seats to party i. (This is usually not an integer.)

The *seat excess* for party i is the difference

$$\Delta_i := s_i - q_i = s_i - p_i n. \tag{2}$$

Note that

$$\sum_{i=1}^{m} \Delta_i = \sum_{i=1}^{m} s_i - n = 0. \tag{3}$$

The *bias* is the mean $\mathbb{E}\,\Delta_i$ of the seat excess.

This assumes that we consider a random instance. Some possibilities:

- A sample of real elections.
  (E.g. Pólya (1918), Balinski and Young (2001).)

The *bias* is the mean $\mathbb{E}\,\Delta_i$ of the seat excess.
This assumes that we consider a random instance. Some possibilities:

- A sample of real elections.
(E.g. Pólya (1918), Balinski and Young (2001).)
- A sample of simulated elections.
(E.g. Balinski and Young (2001).)

The *bias* is the mean $\mathbb{E}\,\Delta_i$ of the seat excess.

This assumes that we consider a random instance. Some possibilities:

- A sample of real elections.
  (E.g. Pólya (1918), Balinski and Young (2001).)
- A sample of simulated elections.
  (E.g. Balinski and Young (2001).)
- Pólya (1918–1919), Pukelsheim and others (2003-): Take $p_1, \ldots, p_m$ random and uniformly distributed but ordered $p_1 \geq p_2 \geq \cdots \geq p_m$.

The *bias* is the mean $\mathbb{E}\,\Delta_i$ of the seat excess.

This assumes that we consider a random instance. Some possibilities:

- A sample of real elections.
  (E.g. Pólya (1918), Balinski and Young (2001).)
- A sample of simulated elections.
  (E.g. Balinski and Young (2001).)
- Pólya (1918–1919), Pukelsheim and others (2003-): Take $p_1, \dots, p_m$ random and uniformly distributed but ordered $p_1 \geq p_2 \geq \dots \geq p_m$.
- My approach: Consider $p_1, \dots, p_m$ as given but let n be random. (Random in $\{1, \dots, N\}$; then let $N \to \infty$.)

Advantages

- Asymptotic results depend on $p_1, \ldots, p_m$ but not on n. Thus assumptions of random n are more robust than assumptions of random p_i.
- More precise information.
  Example: d'Hondt's method, three parties.
  Pólya: The largest party has a bias of $5/12$, the second $-1/12$, the smallest $-4/12$.
  My approach: A party of size p has bias $(3p - 1)/2$.
- Simpler formulas.

Advantages

- Asymptotic results depend on $p_1, \ldots, p_m$ but not on n. Thus assumptions of random n are more robust than assumptions of random p_i.
- More precise information.
  Example: d'Hondt's method, three parties.
  Pólya: The largest party has a bias of $5/12$, the second $-1/12$, the smallest $-4/12$.
  My approach: A party of size p has bias $(3p-1)/2$.
- Simpler formulas.
- Leads to nice and interesting mathematics.

Disadvantages

- In practice, n is not random.
  (But neither does $n \to \infty$ hold.)
- We have to assume that $p_1, \ldots, p_m$ are linearly independent over the rational numbers.
  (But this is implicit when $p_1, \ldots, p_m$ are random, so it is really nothing new.)

For asymtotic results these are not serious problems.

Rounding

$\lfloor x \rfloor$ and $\lceil x \rceil$ means rounding down and up of a real number x.
$\{x\} := x - \lfloor x \rfloor$ is the fractional part of x.

More generally, let α be a real number. The *α-rounding* of a real number x is the integer $[x]_\alpha$ such that

$$x - \alpha \leq [x]_\alpha \leq x - \alpha + 1. \tag{4}$$

Consequently,

$$[x]_\alpha = \lceil x - \alpha \rceil = \lfloor x + 1 - \alpha \rfloor, \tag{5}$$

If $0 \leq \alpha \leq 1$, this means that x is rounded down if its fractional part is less than α and up if its fractional part is greater than α. In particular, $\alpha = \frac{1}{2}$ yields standard rounding, $\alpha = 0$ yields rounding up and $\alpha = 1$ yields rounding down. (But note that we allow also $\alpha < 0$ or $\alpha > 1$, in which case $|x - [x]_\alpha|$ may be greater than 1.)

Election methods

- The *β-stationary divisor method*, or the *divisor method with $d(k) = k + \beta$* (where β is a real number): Let

$$s_i := \left[\frac{v_i}{D}\right]_\beta = \left[\frac{p_i}{D'}\right]_\beta, \tag{6}$$

  where D (or $D' = D/V$) is chosen such that $\sum_{i=1}^m s_i = n$. Examples: $\beta = 1$ (Jefferson, d'Hondt), $\beta = 1/2$ (Webster, Sainte-Laguë), $\beta = 0$ (Adams), $\beta = 2$ (Imperiali).
- The *γ-quota method* (where γ is a real number): Let $Q := V/(n + \gamma)$ and let

$$s_i := \left[\frac{v_i}{Q}\right]_\alpha = [(n+\gamma)p_i]_\alpha, \tag{7}$$

  where α is chosen such that $\sum_{i=1}^m s_i = n$. Examples: $\gamma = 0$ (Hamilton, Hare, method of largest remainder), $\gamma = 1$ (Droop), $\gamma = 2$ (Imperiali).

Asymptotic bias

Theorem

For the β-stationary divisor method:

$$\mathbb{E}\,\Delta_i \to \left(\beta - \tfrac{1}{2}\right)(mp_i - 1). \tag{8}$$

For the γ-quota method:

$$\mathbb{E}\,\Delta_i \to \gamma\left(p_i - \frac{1}{m}\right). \tag{9}$$

The asymptotic bias for a party thus depends only on its size and the number of parties, but not on the sizes of the other parties.

In particular, the bias is 0 for every party when $\beta = 1/2$ (Webster/Sainte-Laguë) or $\gamma = 0$ (Hamilton/Hare). This is well-known with other approaches; our approach confirms this, and shows that the method really is unbiased for a party of any size.

Asymptotic distribution

Theorem

For the β-stationary divisor method:

$$\Delta_i \xrightarrow{\mathrm{d}} \bar{X}_i := (\beta - \tfrac{1}{2})(mp_i - 1) + \widetilde{U}_0 + p_i \sum_{k=1}^{m-2} \widetilde{U}_k. \tag{10}$$

For the γ-quota method:

$$\Delta_i \xrightarrow{\mathrm{d}} Y_i := \gamma\Big(p_i - \frac{1}{m}\Big) + \widetilde{U}_0 + \frac{1}{m} \sum_{k=1}^{m-2} \widetilde{U}_k. \tag{11}$$

Here $\widetilde{U}_k \sim \mathrm{U}(-\tfrac{1}{2}, \tfrac{1}{2})$ are independent.

Asymptotic variance

Corollary

For the β-stationary divisor method:

$$\mathrm{Var}\,\Delta_i \to \frac{1 + (m-2)p_i^2}{12} \tag{12}$$

For the γ-quota method:

$$\mathrm{Var}\,\Delta_i \to \frac{1 + (m-2)/m^2}{12} = \frac{(m+2)(m-1)}{12m^2}. \tag{13}$$

Violating quota?

We say that a seat assignment s_i *satisfies lower quota* if $s_i \geq \lfloor q_i \rfloor$ and *satisfies upper quota* if $s_i \leq \lceil q_i \rceil$; it *satisfies quota* if both holds.

In terms of the seat excess $\Delta_i = s_i - q_i$, the assignment satisfies lower [upper] quota if and only if $\Delta_i > -1$ $[\Delta_i < 1]$, and it satisfies quota if and only if $|\Delta_i| < 1$.

It is well-known that Hamilton/Hare's method always satisfies quota, while Jefferson's and Droop's methods satisfy lower quota and Adams method satisfies upper quota. It is also well-known that Webster/Sainte-Laguë does not always satisfy quota, but that violations are unusual in practice.

The theorem above enable us to calculate the (asymptotic) probabilities that quota is violated for various methods.

Example

Jefferson/d'Hondt's method ($\beta = 1$), and a party i with three times the average size: $p_i = 3/m$.
The bias is 1. It follows by the theorem above that

$$\mathbb{P}(\Delta_i > 1) \to 1/2,$$

so the (asymptotic) probability that the party violates quota is 1/2. For a larger party, the probability is even greater.

Example

The Swedish parliament contains at present 8 parties; two large with 30% of the votes each and 6 small with 5–8% percent each. The seats are in principle distributed by Sainte-Laguë's method ($\beta = 1/2$).

The small parties always satisfy quota. In fact, for Webster/Sainte-Laguë, only parties with $p_i \geq 1/(m-2)$ can violate quota.

For the large parties we have $p_i = 0.3$, and thus

$$\Delta_i \to X_i := \widetilde{U}_0 + 0.3 \sum_{k=1}^{6} \widetilde{U}_k.$$

An integration yields

$$\mathbb{P}(\bar{X}_i \geq 1) = \mathbb{P}(\bar{X}_i \leq -1) = 0.00045.$$

Hence, for each of the two large parties the (asymptotic) probability of violating quota is 0.0009.

The Alabama paradox

(Joint work with Svante Linusson.)

Theorem

The probability that state i suffers from the Alabama paradox when we increase the total number of seats by one equals

$$\frac{1}{m}\,\mathbb{E}\big(S_i^- - S_i^+ - 1\big)_+, \tag{14}$$

where $S_i^+ = \sum_{j:p_j<p_i} I_j^{(i)}$ *and* $S_i^- = \sum_{j:p_j>p_i} I_j^{(i)}$ *with* $I_j^{(i)} \sim \mathrm{Be}(|p_i - p_j|)$ *and* $I_1^{(i)}, \dots, I_m^{(i)}$ *independent.*
If the states are ordered with $p_1 \le p_2 \le \cdots \le p_m$*, this can be written*

$$\frac{1}{m}\sum_{s=0}^{i-1}\sum_{k=2}^{m-i}(-1)^{s+k}\binom{s+k-2}{s}e_s(r_1^{(i)},\dots,r_{i-1}^{(i)})e_k(r_{i+1}^{(i)},\dots,r_m^{(i)}),$$

where $r_j^{(i)} := |p_i - p_j|$ *and* e_k *is the elementary symmetric polynomial.*

Corollary

Suppose that there are three states with relative sizes p_1, p_2, p_3, with $p_1 \leq p_2 \leq p_3$. Then only the smallest state can suffer from the Alabama paradox, and the probability of this is

$$\frac{1}{3}(p_2 - p_1)(p_3 - p_1).$$

The supremum of this probability over all distributions (p_1, p_2, p_3) is $1/12$, and the average is $1/36$.

With (uniform) random poulation sizes, the expected number of occurences of the Alabama paradox is ≈ 0.12324 for large m.

Proofs

Proofs are based on Weyl's theorem: If $W_i = \{np_i\}$ then $W_1, \dots, W_{m-1}$ are (asymptotically) uniformly distributed in $(0, 1)$ and independent.

However, W_m is determined by $W_1, \dots, W_{m-1}$. Moreover, the adjustment of the divisor in divisor methods, or the rounding threshold in quota methods, in order to achieve a fixed house size n introduces some rather complicated dependencies. Nevertheless, after some algebraic manipulations we obtain the results above.

Further applications

Further applications

- The probability that a party or coalition with a small majority of the votes but not a majority of the seats.

Further applications

- The probability that a party or coalition with a small majority of the votes but not a majority of the seats.
- Biproportional methods??