The Probability of The Alabama Paradox

Svante Janson
(joint work with Svante Linusson)

AofA, Będlewo, June 2011

US Constitution (1789)

"Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, three fifths of all other Persons. The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct. The Number of Representatives shall not exceed one for every thirty Thousand, but each State shall have at Least one Representative."
(Article I, Section 2)

"in such Manner as they shall by Law direct"

"in such Manner as they shall by Law direct"

HOW?

"in such Manner as they shall by Law direct"

HOW?

This has been the cause of much political fighting,

"in such Manner as they shall by Law direct"

HOW?

This has been the cause of much political fighting,

but also of some good mathematics.

The same problem arises in many other countries where often seats in parliament are distributed among constituencies according to population.

Moreover, in many countries, the seats are given to parties in proportion to the number of votes.

The same problem arises in many other countries where often seats in parliament are distributed among constituencies according to population.

Moreover, in many countries, the seats are given to parties in proportion to the number of votes.

Many different proportional election methods exist.

Hamilton's method

Suppose that there are m states with populations $P_1, \ldots, P_m$, and n seats to distribute. Let $P = \sum_{i=1}^{m} P_i$ be the total population and let $p_i = P_i/P$ be the relative population of state i, i.e., its proportion of the total population.

Calculate $\mu_i = p_i n$; this is the real number that would give exact proportionality. First round these down and give $\lfloor \mu_i \rfloor$ seats to state i. The sum of these numbers is almost always less than n (the exception is when all μ_i happen to be integers), and the remaining seats are given to the states with largest remainders $\mu_i - \lfloor \mu_i \rfloor$.

In other words, μ_i is rounded up for the states with largest remainders, and the number of states that are rounded up is determined so that the total number of seats becomes n.

As a formula:

$$s_i := \lceil \mu_i - r \rceil = \lceil p_i n - r \rceil$$

for a suitable threshold r.

This is a simple and natural method, and it has been invented and reinvented many times.

It was proposed by Alexander Hamilton (1755–1804) in 1792 for the first apportionment in USA. (Approved by Congress but vetoed by president Washington.)

Proposed again in 1850 when it became law and was used, with some fiddling, for the rest of the century (as Vinton's method).

Used today in many countries, for example Lithuania, Denmark, Germany (before 2009 in federal elections and still in 7 states), ...

Many names:
Hamilton's method,
Hare's method,
Hare/Niemeyer's method,
Method of largest remainder,
...

The Alabama paradox

The method does not bias against small or large states. However, in 1881 it was discovered that this method has a surprising and unwelcome behavior when the total number of seats is changed: It can happen that some state gets *less* representatives when the total number is increased. (The total number was not fixed.)
More precicely, using the population figures from the 1880 census, a total of 299 seats would give 8 to Alabama, but a total number of 300 would give only 7 to Alabama.
This counterintuitive behaviour got the name *Alabama paradox*, and it eventually led to the abolishment of Hamilton's method in favour of others that do not suffer from this defect.
(The same problem was actually observed already 1871, in that case for Rhode Island, but this went largely unnoticed.)

state	pop.	μ_i	seats
A	53	5.30	5
B	33	3.30	3
C	14	1.40	2
sum	100	10.00	10

10 seats

state	pop.	μ_i	seats
A	53	5.83	6
B	33	3.63	4
C	14	1.54	1
sum	100	11.00	11

11 seats

Figure: The Alabama paradox. Numbers in blue are rounded up. (Population figures may be in thousands or millions, for example.)

The Alabama paradox is mathematically not strange, once it has been noticed: Consider three states A, B and C. If we increase n to $n+1$, the number $\mu_i = np_i$ is increased by p_i. If, for example, state A is small and states B and C larger, then μ_{A} increases less that μ_{B} and μ_{C}. Suppose for simplicity that none of these numbers passes an integer, so the integer parts $\lfloor \mu_i \rfloor$ remains the same for n and $n+1$ for all three states; then their remainders $\rho_i = \mu_i - \lfloor \mu_i \rfloor$ increase by p_i, and it may happen that the remainders ρ_{B} and ρ_{C} both are smaller than ρ_{A} when we distribute n seats, but that both become larger than ρ_{A} when we increase n to $n+1$. If furthermore A had the smallest remainder that was rounded up, then the result is that A loses one seat while B and C gain one each. (We assume that no other state interferes.)

Mathmematical problem

So the Alabama paradox may happen, but how likely is it?
Let us assume that the population sizes are given, but choose a *random* number n of seats. *What is the probability that the Alabama paradox occurs if n is increased to $n+1$?*
By choosing a random number n we mean choosing n uniformly at random from $\{1, 2, \ldots, N\}$ for some large integer N, and then taking the limit (assuming that it exists) as $N \to \infty$. Thus, more formally, let $s_i(n)$ be the number of seats state i receives when n seats are distributed. Increase the number of seats n, by one seat at a time, from 1 to N, and let $\nu_i(N)$ be the number of times that state i suffers from the Alabama paradox, i.e., the number of $n < N$ such that $s_i(n+1) < s_i(n)$. If $\nu_i(N)/N$ converges to some value q_i as $N \to \infty$, we say that the limit q_i is the *probability that state i suffers from the Alabama paradox*.

Assumption

We make one mathematical simplification.
A set $\{x_1, \ldots, x_k\}$ of real numbers is *linearly independent over* $\mathbb{Q}$ if there is no relation $a_1 x_1 + \cdots + a_k x_k = 0$ with all coefficients a_i rational and not all 0. (Equivalently, there is no such relation with integer coefficients a_i, not all 0.)
We will assume that the relative population sizes are linearly independent over $\mathbb{Q}$.

Mathematically, this assumption is reasonable, since if we choose $p_1, \ldots, p_m$ at random (uniformly given that their sum is 1), they will almost surely be linearly independent over $\mathbb{Q}$. However, for the practical problem of apportionment, the assumption is clearly unreasonable since the populations P_i are integers and the p_i thus rational numbers. Nevertheless, our formula is a good approximation if the numbers p_i have large denominators and there are no relations

$$a_1 p_1 + \cdots + a_m p_m = 0 \tag{1}$$

with small integers $a_1, \ldots, a_m$.

Theorem

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

$$\frac{1}{m}\,\mathbb{E}\big(S_i^- - S_i^+ - 1\big)_+, \tag{2}$$

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.

Proof

Think of the different states as m runners on a circular track, with state i running at constant speed p_i (laps/time unit). At time n, state i has run a distance np_i, and thus $\lfloor np_i \rfloor$ full laps, so the number of seats that it gets is the number of completed laps, plus an additional seat for each of the states that have come furthest on the next lap; the number of these additional seats is chosen such that the total number of seats is n.

We reformulate this by moving the finishing line; we mark its position by a flag and count laps as runners pass the flag. We place the flag by the runner that got the last additional seat, i.e. the state with the smallest remainder that is rounded up. Then the number of seats a state gets equals its number of laps, for every state.

Let us increase the total number of seats from n to $n + 1$ in two steps.

We first increase time from n to $n + 1$ continuously, letting the runners run, but at the same time we also move the flag, by letting it be carried by a runner, so that the total number of laps stays at n.
This means that if the runner carrying the flag overtakes another, slower, runner, then the flag is passed to the slower runner and both runners keep the same numbers of laps. On the other hand, if the runner carrying the flag is overtaken by a faster runner, then the flag is passed to the faster runner, who gets one lap more, while the former flag-holder loses one lap. (Other overtakings do not affect the flag, nor the number of laps for anyone.)

Finally, at time $n + 1$ we increase the total number of seats by one; this means that the runner carrying the flag throws it to the next runner behind him (her), who gains another lap.

We count positions, at any given time, relative to the flag and say that position 0 is the runner carrying the flag, positions $-1, -2, \ldots$ are the runners behind the flag-carrying runner, and positions $1, 2, 3, \ldots$ the runners in front of him/her. (Since the track is circular, position k and position $k - m$ are the same, but that does not matter as long as we take a little care.)

It is easy to see that when one runner overtakes another, their positions (which differ by 1) are exchanged, while all other positions remain the same; this hold also if one of them carries the flag.

Consider a specific runner, say runner i. The position of i increases by 1 each time i overtakes someone else, and it decreases by 1 each time i is overtaken. Furthermore, it increases by 1 at the final step when the flag is thrown. Thus, if S^+ is the number of runners overtaken by i, and S^- the number of runners overtaking i, during the interval $[n, n+1]$, then the position is increased by $S^+ - S^- + 1$. Since the number of laps is changed only when the position changes between 0 and -1, we see that:

(+) State i gains a seat if $S^+ - S^- + 1 > 0$ and runner i has at time n one of the positions $-1, -2, \dots, -(S^+ - S^- + 1)$.

(−) State i loses a seat if $S^+ - S^- + 1 < 0$ and runner i has at time n one of the positions $0, 1, \dots, |S^+ - S^- + 1| - 1$.

Case (−) is thus when the Alabama paradox occurs for state i.

Let L be the position of runner i relative to the flag at time n, normalized to have $L \in \{0, \dots, m-1\}$. Then the Alabama paradox occurs if and only if

$$S^+ - S^- + 1 < -L. \qquad (3)$$

Let the indicator I_j^+ be 1 if i overtakes j during $[n, n+1]$, and 0 otherwise; similarly, let I_j^- be 1 if i is overtaken by j and 0 otherwise. Then $S^+ = \sum_j I_j^+$ and $S^- = \sum_j I_j^-$. (Note that no runner can overtake another more than once during $[n, n+1]$.)
We let $\{x\} := x - \lfloor x \rfloor \in [0, 1)$. Then

$$I_j^+ = 1 \iff p_i > p_j \text{ and } 0 < \{np_j - np_i\} < p_i - p_j, \tag{4}$$
$$I_j^- = 1 \iff p_i < p_j \text{ and } 1 - (p_j - p_i) < \{np_j - np_i\} < 1. \tag{5}$$

We calculate the probability of $S^+ - S^- + 1 < -L$ by finding the asymptotic joint distribution of L and the fractional parts $\{np_j - np_i\}$, $j \neq i$.
By the formulas above, this gives the asymptotic joint distribution of S^+, S^- and L, and thus the (asymptotic) probability of the Alabama paradox.

Lemma

Suppose that $p_1, \ldots, p_m$ are linearly independent over $\mathbb{Q}$, and let $L_n \in \{0, \ldots, m-1\}$ be the position of runner 1 relative to the flag at time n. Then the sequence of vectors $v_n = (\{n(p_2 - p_1)\}, \ldots, \{n(p_m - p_1)\}, L_n)$, $n \geq 1$, is uniformly distributed on $[0,1)^{m-1} \times \{0, \ldots, m-1\}$.

The proof is based on Weyl's theorem:

If $W_i := \{np_i\}$ then $W_1, \ldots, W_{m-1}$ are (asymptotically) uniformly distributed in $(0,1)$ and independent.

(Note that W_m is determined by $W_1, \ldots, W_{m-1}$.)