THE PROBABILITY OF THE ALABAMA PARADOX

SVANTE JANSON AND SVANTE LINUSSON

Abstract. Hamilton's method is a natural and common method to distribute seats proportionally between states (or parties) in a parliament. In USA it has been abandoned due to some drawbacks, in particular the possibility of the Alabama paradox, but it is still in use in many other countries. In this paper we give, under certain assumptions, a closed formula for the asymptotic probability, as the number of seats tends to infinity, that the Alabama paradox occurs given the vector $p_1, \dots, p_m$ of relative sizes of the states.

From the theorem we deduce a number of consequences. For example it is shown that the expected number of states that will suffer from the Alabama paradox is asymptotically bounded above by $1/e$ and on average approximately 0.123.

1. Introduction and main result

Proportional representation is desired in various circumstances. One common case is in elections in many countries, where each party is awarded a number of seats in parliament proportional to the number of votes. Since the number of seats has to be an integer, it becomes a mathematical problem to choose these integers in a way that approximates exact proportionality, and a number of different methods are in use.

In the United States, elections are done differently (with single-member constituences), but the same mathematical problem exists for apportionment to the House of Representatives. By the Constitution, each state has a number of representatives that is proportional to its population. However, the Constitution does not specify by which method the numbers are to be determined. (The numbers are determined by Congress every tenth year, after a new census.) Therefore, the choice of method has been subject to much debate since 1791, see Balinski and Young [1] for a detailed history. (In 1941, a specific method was chosen by law to be used not only for that apportionment but also for all coming ones. This has eliminated the need

Date: 12 April, 2011; revised 14 December, 2011.

2000 *Mathematics Subject Classification.* 60C05; 91B12.

Key words and phrases. Alabama paradox; election methods; apportionment; proportional allocation; Hamilton's method; method of largest remainder.

Svante Linusson is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

for debates and new choices every ten years, so there is now much less debate. For the current method, see [1], which also discusses why the method is slightly biased and could be improved.)

The problem was further complicated by the fact that the Constitution does not specify the total number of representatives. Thus, when discussing apportionment, the Congress discussed not only different methods but also different sizes of the House. During the second half of the 19th century, the favourite method was *Hamilton's method* (also called *method of largest remainder*), which can be described as follows. (The method was proposed by Alexander Hamilton in 1792 for the first apportionment; it was then approved by Congress but vetoed by president Washington. The method was proposed again by Samuel Vinton in 1850 when it became law and was used, with some fiddling, for the rest of the century; see [1] for details.)

Hamilton's method. *Suppose that there are m states with populations $P_1, \dots, P_m$, 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.*

The method is simple and intuitive, and it 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. 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, see [1, p. 39] for details. 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 [1, p. 38].)

Remark 1.1. The method is still used in parliamentary elections in several countries, either to distribute seats among multi-member consituencies (e.g. Sweden), or to distribute seats among the parties (e.g. Denmark). There the number of seats is fixed in advance, so the Alabama paradox is not an obvious problem. However, the paradox may surface and give strange behaviour in combination with other rules, and in some election systems where the method is used a party that gains a vote can, in exceptional

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

situations, actually lose a seat in parliament, see [3]. Germany used earlier Hamilton's method (there called Hare–Niemeyer's method) for federal elections but changed in 2008 for this reason (although the problem partly remains for other reasons); the method is still used in several German states, however.

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 C is small and states A and B larger, then μ_{C} increases less that μ_{A} and μ_{B}. 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_{A} and ρ_{B} both are smaller than ρ_{C} when we distribute n seats, but that both become larger than ρ_{C} when we increase n to $n+1$. If furthermore C had the smallest remainder that was rounded up, then the result is that C loses one seat while A and B gain one each. (We assume that no other state interferes.) A simple numerical example is given in Figure 1.

So the Alabama paradox certainly may happen, and it has occurred, but how likely is it?

Of course, the description so far is purely deterministic (except for the necessity to draw lots sometimes when there is a tie); the Alabama paradox either occurs or not for given parameters. Let us, however, 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 $\{1, 2, \dots, N\}$ 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.* (This approach, to consider given sizes but a random number of seats, is used also in [5] where some other properties of election methods are studied.)

In order to calculate this limit (and show that it exists), we will make one mathematical simplification. Recall that a set $\{x_1,\dots,x_k\}$ of real numbers is *linearly independent over* $\mathbb{Q}$ if there is no relation $a_1x_1+\dots+a_kx_k=0$ 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 $\mathbb{Q}$.

Remark 1.2. Mathematically, this assumption is reasonable, since if we choose $p_1,\dots,p_m$ 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, the formula below is a good approximation if the numbers p_i have large denominators and there are no relations

$$a_1p_1+\dots+a_mp_m=0 \tag{1.1}$$

with small integers $a_1,\dots,a_m$. More precisely, it will be shown in Section 7.2 that for any $\varepsilon>0$, there is an $A=A(m,\varepsilon)$ such that the value q_i in (1.2) or (1.3) differs from the exact probability by less than ε for every distribution $(p_i)_1^m$ for which there is no such relation (1.1) with integers a_i and $\sum_i |a_i|\le A$; we omit the details.

We leave it as an open problem to extend the result below and find exact formulas for all $p_1,\dots,p_m$, and in particular for rational $p_1,\dots,p_m$. (If $p_1,\dots,p_m$ are rational, then the sequence $s_i(n+1)-s_i(n)$ will be periodic, so the limit q_i certainly exists; the existence in general is shown in Section 7.2.) Note that some modifications are required for rational $p_1,\dots,p_m$. For example with three states and $p_1=p_2=2/5$, $p_3=1/5$, it is easily seen that the Alabama paradox never occurs, so the probability is 0 for all three states. See also Proposition 7.4, were it is shown that the expected number of states suffering from the paradox could be arbitrarily close to 1.

We use the standard notations $x_+ := \max\{x,0\}$ and $x_- := (-x)_+ = -\min\{x,0\}$, noting $x=x_+-x_-$ and $|x|=x_++x_-$. Let $e_k(x_1,\dots,x_n)$ denote the *elementary symmetric polynomial* of degree k in n variables, i.e. $e_k(x_1,\dots,x_n) := \sum_{1\le i_1<\dots<i_k\le n}\prod_{j=1}^k x_{i_j}$. With $\mathrm{Be}(p)$ we mean the Bernoulli distribution; thus $I\sim\mathrm{Be}(p)$ if $\mathbb{P}(I=0)=1-p$ and $\mathbb{P}(I=1)=p$. We let $p_{(1)}\ge\dots\ge p_{(m)}$ be the population vector $p_1,\dots,p_m$ rearranged in increasing order, and let $q_{(1)},\dots,q_{(m)}$ be the corresponding probabilities of the Alabama paradox, which by Corollary 2.3 below is the vector $q_1,\dots,q_m$ rearranged in increasing order. For clarity, we will use the notation $p_{(i)}$ and $q_{(i)}$ whenever we consider the states in increasing order, and p_i, q_i only when the order is irrelevant.

Theorem 1.3. *Suppose that m states have relative sizes $p_1,\dots,p_m$, with $\sum_{i=1}^m p_i=1$, and assume that $p_1,\dots,p_m$ are linearly independent over $\mathbb{Q}$.*

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

$$q_i := \frac{1}{m}\,\mathbb{E}\bigl(S_i^- - S_i^+ - 1\bigr)_+, \tag{1.2}$$

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. More explicitly, if the states are ordered with $p_{(1)} \geq \dots \geq p_{(m)}$, this can be written

$$q_{(i)} = \frac{1}{m}\sum_{s=0}^{m-i}\sum_{k=2}^{i-1}(-1)^{s+k}\binom{s+k-2}{s}e_k(\bar{r}_1^{(i)},\dots,\bar{r}_{i-1}^{(i)})e_s(\bar{r}_{i+1}^{(i)},\dots,\bar{r}_m^{(i)}), \tag{1.3}$$

where $\bar{r}_j^{(i)} := |p_{(i)} - p_{(j)}|$.

In other words, each $I_j^{(i)} \in \{0,1\}$ with $\mathbb{P}(I_j^{(i)} = 1) = |p_i - p_j|$.

Remark 1.4. If we do not order the states, then equation (1.3) can equivalently be written as

$$q_i = \frac{1}{m}\sum_{s=0}^{m-3}\sum_{k=2}^{m-s-1}(-1)^{s+k}\binom{s+k-2}{s}e_s\bigl(r_{1+}^{(i)},\dots,r_{m+}^{(i)}\bigr)e_k\bigl(r_{1-}^{(i)},\dots,r_{m-}^{(i)}\bigr), \tag{1.4}$$

where $r_{j\pm}^{(i)} := (p_i - p_j)_\pm$. Since (1.4) is symmetric under permutations of $p_1,\dots,p_m$, we may assume $p_{(1)} \geq \dots \geq p_{(m)}$. In this case, $r_{j+}^{(i)} = 0$ for $j \leq i$ and $r_{j-}^{(i)} = 0$ for $j \geq i$, and it is easily seen that the sums in (1.4) and (1.3) are equal.

The proof of the theorem is given in Section 4. We give first several consequences of the main theorem in Sections 2–3; the proofs of these results are given in Sections 5–6.

For simplicity, we have here considered one state at a time. It may happen that the Alabama paradox occurs for two (or more) states at the same time, although this is less likely, see Section 7.1.

2. Further results

In the case of three states, Theorem 1.3 yields the following simple formula.

Corollary 2.1. *Suppose that there are three states with relative sizes $p_{(1)} \geq p_{(2)} \geq p_{(3)}$ with $p_{(1)} + p_{(2)} + p_{(3)} = 1$, and assume that $p_{(1)}, p_{(2)}, p_{(3)}$ are linearly independent over $\mathbb{Q}$. Then only the smallest state can suffer from the Alabama paradox, and the probability of this is $\frac{1}{3}(p_{(1)} - p_{(3)})(p_{(2)} - p_{(3)})$.*

The supremum of this probability over all distributions $(p_{(1)}, p_{(2)}, p_{(3)})$ is $1/12$, which is never attained but is approached in the extreme case when $p_{(3)}$ is very small and $p_{(1)}$ and $p_{(2)}$ both are close to $1/2$.

In general, the Alabama paradox can affect any state except the two largest, but it is much more likely to affect small states.

Remark 2.2. Note however that Hamilton's method is unbiased. On the average, state i increases its number of seats by p_i each time n is increased, so if it sometimes suffers from the Alabama paradox and its representation decreases with frequency q_i, this must be compensated by a frequency p_i+q_i of the times when the number of seats increases.

Corollary 2.3. *In addition to the assumptions of Theorem 1.3, assume that* $p_{(1)} \geq \cdots \geq p_{(m)}$. *Then* $q_{(m)} \geq q_{(m-1)} \geq \cdots \geq q_{(3)} \geq q_{(2)} = q_{(1)} = 0$. *Moreover, the largest probability is*

$$q_{(m)} = \frac{1}{m} \prod_{j=1}^{m-1} \bigl(1 - (p_{(j)} - p_{(m)})\bigr) - p_{(m)}. \tag{2.1}$$

We have, for any i, the inequalities

$$\frac{1}{m}\Bigl(e^{-1} - mp_{(i)} - \frac{1}{2}\sum_j p_j^2\Bigr) \leq q_{(i)} < \frac{1}{m}\Bigl(1 - \frac{1}{m-1}\Bigr)^{m-1} < \frac{1}{m}e^{-1}. \tag{2.2}$$

If $p_{(m)} \to 0$ and all other $p_{(j)} \to 1/(m-1)$, then (2.1) shows that $q_{(m)} \to \frac{1}{m}\bigl(1 - \frac{1}{m-1}\bigr)^{m-1}$, so this is, for a given m, the least upper bound on the probability of the Alabama paradox for a specific state in the linearly independent case (but only in that case, see Section 7.2), which for large m approaches $1/me$.

Corollary 2.4. *Under the assumptions of Theorem 1.3, the expected number of states suffering from the Alabama paradox each time the number of seats is increased is less than $1/e$. This bound is approached if we let $m \to \infty$ and suppose that $m - o(m)$ of the states are very small, with $p_i = o(1/m)$, and that the remaining states are medium-size with $p_i = o(1)$.*

In this extremal case, the paradox is thus very common. It can be even more common in the rational case, see Proposition 7.4. See also Example 7.2, where we show that the probability of at least one state suffering the Alabama paradox in this case converges to $1 - 2/e \approx 0.264$. We conjecture that this is the upper bound of the probability that at least one state suffers the paradox (under the assumption of Theorem 1.3); note that Corollary 2.4 shows that the probability is always less than $e^{-1} \approx 0.368$.

The exact formula in (1.3) is unwieldy when m is large; it may then be attractive to use (1.2) with a Poisson approximation of $S_i^{\pm}$:

Corollary 2.5. *Under the assumptions of Theorem 1.3, let (for a given i)* $\lambda^+ := \sum_{j:p_j<p_i}(p_i - p_j)$ *and* $\lambda^- := \sum_{j:p_j>p_i}(p_j - p_i)$, *Let further* $\widehat{S}^+ \sim \mathrm{Po}(\lambda^+)$ *and* $\widehat{S}^- \sim \mathrm{Po}(\lambda^-)$ *be independent Poisson random variables, and define*

$$\hat{q}_i := \frac{1}{m}\,\mathbb{E}\bigl(\widehat{S}^- - \widehat{S}^+ - 1\bigr)_+ \tag{2.3}$$

$$= \frac{1}{m} \sum_{j \geq k+2} (j-k-1) \frac{(\lambda^-)^j (\lambda^+)^k}{j!\, k!} e^{-\lambda^- - \lambda^+} \tag{2.4}$$

$$= \frac{1}{m} \sum_{j \leq k} (k+1-j) \frac{(\lambda^-)^j (\lambda^+)^k}{j!\, k!} e^{-\lambda^- - \lambda^+} - p_i. \tag{2.5}$$

Then

$$|q_i - \hat{q}_i| \leq \frac{1}{m} \sum_{j=1}^{m} (p_j - p_i)^2. \tag{2.6}$$

3. Average probability of the Alabama paradox

We have so far considered the probability of the Alabama paradox for given relative population sizes $p_1, \dots, p_m$. Let us now instead fix $m \geq 3$ and consider the average probability over all population distributions. In other words, in this section we let $(p_1, \dots, p_m)$ be random and uniformly distributed over the simplex $\Delta^m := \{(p_1, \dots, p_m) \in [0,1]^m : \sum_i p_i = 1\}$, and take the expectation. Note that then $p_1, \dots, p_m$ are linearly independent over $\mathbb{Q}$ a.s., so we may in the sequel assume that Theorem 1.3 and its corollaries apply.

As above, $p_{(1)} \geq \dots \geq p_{(m)}$ denotes the population vector $p_1, \dots, p_m$ rearranged in increasing order; note that $(p_{(1)}, \dots, p_{(m)})$ is uniformly distributed over the subset $p_{(1)} \geq \dots \geq p_{(m)}$ of the simplex Δ^m. Similarly, $q_{(1)} \leq \dots \leq q_{(m)}$ are the corresponding probabilities of the Alabama paradox, which by Corollary 2.3 are $q_1, \dots, q_m$ rearranged in increasing order. In particular, $q_{(m)}$ is the probability of the Alabama paradox for the smallest state. Note that $q_{(1)} = q_{(2)} = 0$ but a.s. $q_{(k)} > 0$ for $k \geq 3$. Since q_i and $q_{(i)}$ are functions of $p_1, \dots, p_m$, they too now are random variables, and we may ask for their expectations, or other properties of their distributions. We use the notations $\xrightarrow{\mathrm{d}}$ and $\xrightarrow{\mathrm{p}}$ for convergence in distribution and probability, respectively, always as $m \to \infty$.

In the case $m = 3$, the average of the formula in Corollary 2.1 over all $(p_{(1)}, p_{(2)}, p_{(3)})$ is easily found by integration, for example by the substitution $p_{(1)} = 1 - p_{(3)} - p_{(2)}$ and integrating over $(p_{(2)}, p_{(3)})$ with the conditions $0 < p_{(3)} < \frac{1}{3}$ and $p_{(3)} < p_{(2)} < (1 - p_{(3)})/2$; a calculation yields the probability $\mathbb{E}\, q_{(3)}$ for the Alabama paradox for three states of random sizes as $1/36$.

We extend this to larger m. First we consider only the smallest state, which is most likely to experience the paradox. Recall the notation $m^{\overline{k}} := m(m+1)\dots(m+k-1)$ for the rising factorial.

Theorem 3.1. *With uniformly random relative population sizes, the expected probability $\mathbb{E}\, q_{(m)}$ that the smallest state among m states will suffer from the Alabama paradox is*

$$\mathbb{E}\, q_{(m)} = \frac{1}{m} \sum_{k=0}^{m-1} (-1)^k \frac{\binom{m-1}{k}}{m^{\overline{k}}} - \frac{1}{m^2} = \frac{1}{m} \sum_{k=2}^{m-1} (-1)^k \frac{\binom{m-1}{k}}{m^{\overline{k}}} \tag{3.1}$$

Hence, as $m \to \infty$,

$$\mathbb{E}\, q_{(m)} = \frac{e^{-1}}{m} - \frac{1}{m^2} + O\left(\frac{1}{m^3}\right). \tag{3.2}$$

Furthermore, $mq_{(m)} \xrightarrow{\mathrm{p}} e^{-1}$ as $m \to \infty$.

Hence, for large m, the expected value of the largest probability $q_{(m)}$ is asymptotically equal to the upper bound e^{-1}/m given in Corollary 2.3, and, furthermore, $q_{(m)}$ is close to e^{-1}/m for most $p_1, \dots, p_m$.

Remark 3.2. The sum in (3.1) is a hypergeometric sum and the result can be written

$$\mathbb{E}\, q_{(m)} = \frac{1}{m}{}_1F_1(1-m; m; 1) - \frac{1}{m^2} \tag{3.3}$$

with a confluent hypergeometric function ${}_1F_1$ (in this case a polynomial).

For small m we have the following table.

m	2	3	4	5	6	7	8	9
$\mathbb{E}\, q_{(m)}$	0	$\frac{1}{36}$	$\frac{17}{480}$	$\frac{61}{1680}$	$\frac{907}{25920}$	$\frac{153709}{4656960}$	$\frac{855383}{27675648}$	$\frac{134964353}{4670265600}$

For a fixed (or random) state, we obtain a more complicated formula.

Theorem 3.3. *With uniformly random relative population sizes the average probability $\mathbb{E}\, q_m$ that a given state among m states will suffer from the Alabama paradox is*

$$\mathbb{E}\, q_m = \frac{1}{m}\sum_{s=0}^{m-3}\sum_{k=2}^{m-s-1}\sum_{i=0}^{s}\sum_{j=0}^{s-i}(-1)^{k+i+j}\binom{s+k-2}{s}$$
$$\times \frac{(m-1)!^2}{k!\,i!\,(s-i-j)!\,(m-1-k-s)!\,(m-1+k+s)!}(i+k+1)^{-j-1}.$$

For small m we have the following table.

m	2	3	4	5	6	7	8
$\mathbb{E}\, q_m$	0	$\frac{1}{108}$	$\frac{17}{1440}$	$\frac{523}{43200}$	$\frac{2287039}{195955200}$	$\frac{100704757}{9144576000}$	$\frac{404675341849}{39230231040000}$

The average probability $\mathbb{E}\, q_m$ in Theorem 3.3 is, of course, always smaller than the average of the largest probability $\mathbb{E}\, q_{(m)}$ in Theorem 3.1. It is somewhat surprising that it is not much smaller, the ratio is close to $1/3$, as is shown for some small m by the following table (with rounded values computed by Maple):

m	3	10	20	30	50	100
$\mathbb{E}\, q_m / \mathbb{E}\, q_{(m)}$	0.33333	0.33392	0.33441	0.33457	0.33474	0.33487

Indeed, this persists for large m, and $\mathbb{E}\, q_m$ is really of order $1/m$, since Theorem 3.3 implies the following asymptotic formula by dominated convergence:

Corollary 3.4. *With uniformly random relative population sizes, the expected number of occurrences of the Alabama paradox among all m states is $m\,\mathbb{E}\,q_m$ which as $m \to \infty$ has the limit*

$$m\,\mathbb{E}\,q_m \to b := \sum_{s=0}^{m-3}\sum_{k=2}^{m-s-1}\sum_{i=0}^{s}\sum_{j=0}^{s-i}(-1)^{k+i+j}\binom{s+k-2}{s}\frac{(i+k+1)^{-j-1}}{k!\,i!\,(s-i-j)!}.$$

Thus, $\mathbb{E}\,q_m \sim b/m$. We do not know any better closed form for b, but numerically we obtain from Maple $b \approx 0.12324$ and thus, using Theorem 3.1, $\mathbb{E}\,q_m/\,\mathbb{E}\,q_{(m)} \to be \approx 0.33501$, in accordance with the table above. (If Maple is correct, the limit is not exactly $1/3$, but quite close.)

The formula for b in Corollary 3.4 as an alternating quadruple sum is not very illuminating, and it is not even easy to see that $b > 0$ from it, but that at least follows from the alternative representation in the next theorem, which adds more information on the asymptotic relation between size and probability for the Alabama paradox for random populations.

Theorem 3.5. *Define, for any $\lambda^-, \lambda^+ \geq 0$, with $\widehat{S}^\pm \sim \mathrm{Po}(\lambda^\pm)$ independent, similarly to* (2.3)–(2.5),

$$\begin{aligned}\Phi(\lambda^-,\lambda^+) &:= \mathbb{E}\big(\widehat{S}^- - \widehat{S}^+ - 1\big)_+ &(3.4)\\ &= \sum_{j\geq k+2}(j-k-1)\frac{(\lambda^-)^j(\lambda^+)^k}{j!\,k!}e^{-\lambda^- - \lambda^+} &(3.5)\\ &= \sum_{j\leq k}(k+1-j)\frac{(\lambda^-)^j(\lambda^+)^k}{j!\,k!}e^{-\lambda^- - \lambda^+} + \lambda^- - \lambda^+ - 1, &(3.6)\end{aligned}$$

and $\Psi(x) := \Phi\big(e^{-x}, e^{-x} - 1 + x\big)$. Then, as $m \to \infty$, with random population sizes,

$$(mp_m, mq_m) \xrightarrow{\mathrm{d}} \big(T, \Psi(T)\big), \tag{3.7}$$

where $T \sim \mathrm{Exp}(1)$. In particular,

$$\begin{aligned}m\,\mathbb{E}\,q_m &\to b = \mathbb{E}\,\Psi(T)\\ &= \sum_{j\geq k+2}(j-k-1)\int_0^\infty \frac{e^{-jx}(e^{-x}-1+x)^k}{j!\,k!}e^{-2e^{-x}-2x+1}\,\mathrm{d}x.\end{aligned} \tag{3.8}$$

By Theorem 3.5 and its proof, we can further say that, assuming random, uniformly distributed populations, a state with $p_i = x/m$ has probability $q_i \approx \Psi(x)/m$ of suffering from the Alabama paradox. Note that the extreme case in Theorem 3.1 can be seen (informally) as the limiting case $x = 0$, since we have $mp_{(m)} \xrightarrow{\mathrm{p}} 0$ and $mq_{(m)} \to e^{-1}$, and indeed $\Psi(0) = \Phi(1,0) = e^{-1}$, e.g. by (3.6).

We plot the function Ψ in Figure 2.

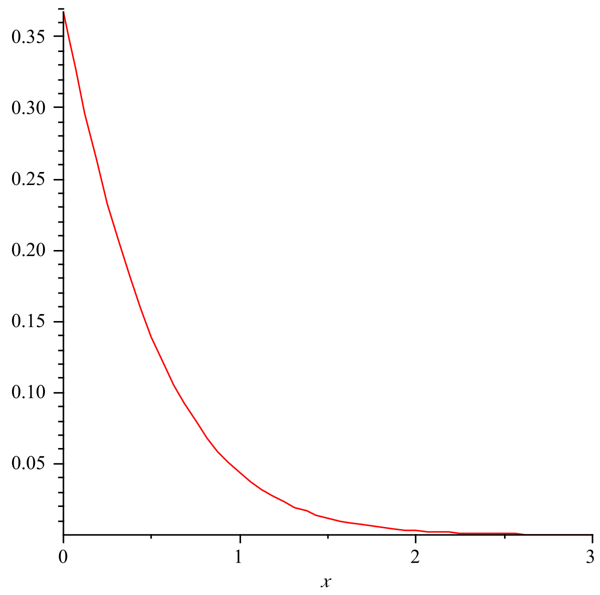

FIGURE 2. The function Ψ in Theorem 3.5.

4. Proof of Theorem 1.3

We analyse the process of successive distributions of seats as follows.

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. (The assumption that $p_1, \dots, p_m$ are linearly independent over $\mathbb{Q}$ implies that ties cannot occur, so we do not have to worry at all about ties in this proof.)

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 $-1, -2, \dots$ are the runners behind the flag-carrying runner, and positions $1, 2, 3, \dots$ 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

$$S^+ - S^- + 1 < -L. \tag{4.1}$$

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 $\{x\} := x - \lfloor x \rfloor \in [0, 1)$ denote the fractional part of a real number x. (We will also use $\{\}$ to denote sets; the meaning should be clear from the context.) Then

$$I_j^+ = 1 \iff p_i > p_j \text{ and } 0 < \{np_j - np_i\} < p_i - p_j, \tag{4.2}$$

$$I_j^- = 1 \iff p_i < p_j \text{ and } 1 - (p_j - p_i) < \{np_j - np_i\} < 1. \tag{4.3}$$

We calculate the probability of (4.1) by finding the asymptotic joint distribution of L and the fractional parts $\{np_j - np_i\}$, $j \neq i$, where again we choose n uniformly at random with $1 \le n \le N$, and then let $N \to \infty$. By the formulas above, this gives the asymptotic joint distribution of S^+, S^- and L, and thus the (asymptotic) probability of (4.1).

We say that an infinite sequence $(v_n)_{n\geq 1} \in [0,1)^{m-1} \times \{0,\dots,m-1\}$ is *uniformly distributed* if the empirical distributions $N^{-1}\sum_{n=1}^{N}\delta_{v_n}$ converge to the uniform distribution as $N\to\infty$, where δ_{v_n} denotes the Dirac measure. This means that if $A \subseteq [0,1)^{m-1}$ with $\lambda(\partial A)=0$ and $k\in\{0,\dots,m-1\}$, then $\#\{n\leq N : v_n \in A\times\{k\}\}/N \to \lambda(A)/m$. (Here λ is the usual Lebesgue measure.) This is a simple extension of the standard notion of uniform distribution for a sequence in $[0,1)^{m-1}$. We claim the following. (For notational convenience, we state the case $i=1$ only.)

Lemma 4.1. *Suppose that $p_1,\dots,p_m$ are linearly independent over $\mathbb{Q}$, and let $L_n \in \{0,\dots,m-1\}$ be the position of runner 1 relative to the flag at time n. Then the sequence of vectors $v_n = (\{n(p_2-p_1)\},\dots,\{n(p_m-p_1)\},L_n)$, $n\geq 1$, is uniformly distributed on $[0,1)^{m-1}\times\{0,\dots,m-1\}$.*

We postpone the proof and first complete the proof of (1.2).

By (4.2) and (4.3), for each j, at most one of I_j^+ and I_j^- is non-zero, depending on whether $p_j<p_i$ or $p_j>p_i$. We simplify the notation by letting $I_j = I_j^+ + I_j^-$; thus $S^+ = \sum_{j:p_j<p_i} I_j$ and $S^- = \sum_{j:p_j>p_i} I_j$.

For a given N, these are random variables, and we have, letting $\mathbb{P}_N$ denote the probability when n is uniformly chosen with $n\leq N$,

$$\mathbb{P}_N(S^+ - S^- + 1 < -L) = \sum_{\ell=0}^{m-1}\mathbb{P}_N(L=\ell \text{ and } S^- - S^+ - 1 > \ell). \tag{4.4}$$

As $N\to\infty$, Lemma 4.1 and (4.2)–(4.3) show that the distribution of $L=L_n$ converges to the uniform distribution on $\{0,\dots,m-1\}$ and the distribution of I_j converges to $\mathrm{Be}(|p_i-p_j|)$ for all $j\neq i$; moreover, these are asymptotically independent. Hence, (4.4) yields

$$\begin{aligned}\mathbb{P}_N(S^+ - S^- + 1 < -L) &\to \sum_{\ell=0}^{m-1}\frac{1}{m}\,\mathbb{P}(S_i^- - S_i^+ - 1 > \ell)\\ &= \frac{1}{m}\,\mathbb{E}\big(S_i^- - S_i^+ - 1\big)_+,\end{aligned}$$

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 $I_j^{(i)} \sim \mathrm{Be}(|p_i-p_j|)$ independent. This is the result stated in (1.2).

We proceed to show Lemma 4.1. First recall a well-known result by Weyl. (The standard proof is by showing that the Fourier transform (characteristic function) $\frac{1}{N}\sum_{n=1}^{N}\exp\big(2\pi\mathrm{i}\sum_{j=1}^{k} n_j\{ny_j\}\big) \to 0$, as $N\to\infty$, for any fixed integers $n_1,\dots,n_k$, not all 0, see for example [4, Exercises 3.4.2–3].)

Lemma 4.2 (Weyl)**.** *Suppose that $y_1,\dots,y_k$ and 1 are linearly independent over $\mathbb{Q}$. Then the sequence of vectors $(\{ny_1\},\dots,\{ny_k\}) \in [0,1)^k$ is uniformly distributed in $[0,1)^k$.* $\square$

We will need the following extension. Let $\mathrm{Mod}_m(x) = m\{x/m\}$; this is the remainder when x is divided by m. Thus, if r is an integer, then

$\mathrm{Mod}_m(r)$ is the unique integer in $\{0,\dots,m-1\}$ such that $\mathrm{Mod}_m(r) \equiv r \pmod{m}$.

Lemma 4.3. *Suppose that $y_1,\dots,y_k$ and 1 are linearly independent over $\mathbb{Q}$. Let $\ell_n = \mathrm{Mod}_m\big(\big(\sum_{j=1}^k \lfloor ny_j\rfloor\big) - n\big) \in \{0,\dots,m-1\}$. Then the sequence of vectors $(\{ny_1\},\dots,\{ny_k\},\ell_n) \in [0,1)^k \times \{0,\dots,m-1\}$ is uniformly distributed in $[0,1)^k \times \{0,\dots,m-1\}$.*

Proof. Let $z_j = y_j/m$ and $w_n = (w_{n1},\dots,w_{nk})$ with $w_{nj} = \{nz_j\}$, $j = 1,\dots,k$. Then $z_1,\dots,z_k$ and 1 are linearly independent over $\mathbb{Q}$, and thus Lemma 4.2 (Weyl's theorem) shows that the sequence $(w_n)_{n\ge1}$ is uniformly distributed in $[0,1)^k$. Further,

$$ny_j - mw_{nj} = mnz_j - m\{nz_j\} = m\lfloor nz_j\rfloor \equiv 0 \pmod{m}.$$

Hence,

$$\{ny_j\} = \{mw_{nj}\} \qquad \text{and} \qquad \lfloor ny_j\rfloor \equiv \lfloor mw_{nj}\rfloor \pmod{m}. \tag{4.5}$$

Thus, $\mathrm{Mod}_m(\lfloor ny_j\rfloor) = \lfloor mw_{nj}\rfloor$.

Let $\tilde\ell_{nj} = \mathrm{Mod}_m(\lfloor ny_j\rfloor) = \lfloor mw_{nj}\rfloor$. If a sequence (u_n) is uniformly distributed in $[0,1)$, then (mu_n) is uniformly distributed in $[0,m)$ and the vectors $(\{mu_n\},\lfloor mu_n\rfloor)$ are uniformly distributed in $[0,1)\times\{0,\dots,m-1\}$. Using this argument in each coordinate, the fact that (w_n) is uniformly distributed in $[0,1)^k$ implies that the sequence of vectors $(\{ny_j\},\tilde\ell_{nj})_{j=1}^k = (\{mw_{nj}\},\lfloor mw_{nj}\rfloor)_{j=1}^k$ is uniformly distributed in $[0,1)^k \times \{0,\dots,m-1\}^k$.

Let $\tilde\ell_n = \mathrm{Mod}_m\big(\sum_{j=1}^k \lfloor ny_j\rfloor\big)$. Then $\tilde\ell_n = \mathrm{Mod}_m\big(\sum_{j=1}^k \tilde\ell_{nj}\big)$, and it follows that the sequence of vectors $(\{ny_1\},\dots,\{ny_k\},\tilde\ell_n)$ is uniformly distributed in $[0,1)^k\times\{0,\dots,m-1\}$.

This is almost what we claim. To complete the proof, we consider a subsequence of the form $n = m\nu + n_0$, $\nu \ge 1$. Weyl's theorem holds for such subsequences too, as a consequence of the version in Lemma 4.2 since $\{(m\nu+n_0)p_j\} = \{\nu(mp_j) + n_0p_j\}$ where $mp_1,\dots,mp_k$ and 1 are linearly independent over $\mathbb{Q}$, and the constant shift by n_0p_j does not affect the uniform distribution. Consequently, the argument above shows that $(\{ny_1\},\dots,\{ny_k\},\tilde\ell_n)$ is uniformly distributed for each such subsequence. But along the subsequence, $\ell_n = \mathrm{Mod}_m(\tilde\ell_n - n_0)$, so $(\{ny_1\},\dots,\{ny_k\},\ell_n)$ is uniformly distributed for each such subsequence, and thus for the entire sequence. □

Proof of Lemma 4.1. Suppose that runner 1 carries the flag at time n, i.e., $L_n = 0$. Then state 1 gets an additional seat, i.e., its number of seats is rounded up to $\lceil np_1\rceil$, and state j gets

$$\lceil np_j - \{np_1\}\rceil = \lceil np_j - np_1\rceil + \lfloor np_1\rfloor = \lfloor np_j - np_1\rfloor + \lceil np_1\rceil$$

seats. Since the total number of seats is n, we have, still in the case $L_n = 0$,

$$n = \sum_{j=2}^{m} \lfloor np_j - np_1 \rfloor + m\lceil np_1 \rceil \equiv \sum_{j=2}^{m} \lfloor np_j - np_1 \rfloor \pmod m.$$

In general, there are L_n additional states whose numbers of seats are rounded up (or $-L_n$ fewer, if $L_n < 0$); thus

$$n \equiv L_n + \sum_{j=2}^{m} \lfloor np_j - np_1 \rfloor \pmod m$$

and

$$L_n \equiv n - \sum_{j=2}^{m} \lfloor n(p_j - p_1) \rfloor \pmod m. \tag{4.6}$$

Let $y_j = p_j - p_1$, $j = 2, \dots, m$. Since $p_1, \dots, p_m$ are linearly independent over $\mathbb{Q}$, it is easily seen that $y_2, \dots, y_m$ and $\sum_1^m p_j = 1$ also are linearly independent over $\mathbb{Q}$. (This can be seen as a change of basis, using a non-singular integer matrix, in a vector space of dimension m over $\mathbb{Q}$.) Thus Lemma 4.1 follows from Lemma 4.3 (with $k = m - 1$, after renumbering $y_2, \dots, y_m$), since $L_n \equiv -\ell_n \pmod m$ by (4.6). $\square$

This completes the proof of (1.2). We proceed to show that this can be evaluated as (1.3).

We may assume that the states are ordered by size, $p_1 \geq \cdots \geq p_m$; thus $p_i = p_{(i)}$. Consider the i:th largest state. Let $X_- \subseteq [i-1]$ and $X_+ \subseteq \{i+1, \dots, m\}$, where as usual $[n] := \{1, \dots, n\}$. We think of X_- and X_+ as the indices j for which $I_j^{(i)} = 1$. Let $P_=(X_-, X_+)$ be the probability that $\{j : I_j^{(i)} = 1\} = X_- \cup X_+$. Then clearly, for simplicity writing $\bar{r}_j := \bar{r}_j^{(i)} = |p_i - p_j|$,

$$P_=(X_-, X_+) = \prod_{j \in X_- \cup X_+} \bar{r}_j \prod_{j \in [m] \setminus (X_- \cup X_+ \cup \{i\})} (1 - \bar{r}_j).$$

The formula (1.2) for the probability that the i:th largest state suffers from the Alabama paradox is thus

$$mq_{(i)} = \sum_{X_- \subseteq [i-1], X_+ \subseteq \{i+1, \dots, m\}} (|X_-| - |X_+| - 1) \cdot \prod_{j \in X_- \cup X_+} \bar{r}_j \prod_{j \in [m] \setminus (X_- \cup X_+ \cup \{i\})} (1 - \bar{r}_j),$$

where the sum runs over all pairs of subsets such that $|X_-| \geq |X_+| + 2$.

Now, for any $s \geq 0$, $k \geq 2$ and any monomial $\prod_{\mu=1}^{k} \bar{r}_{j_\mu} \cdot \prod_{\nu=1}^{s} \bar{r}_{l_\nu}$, where $1 \leq j_1 < \cdots < j_k \leq i-1$ and $i+1 \leq l_1 < \cdots < l_s \leq m$, we get the coefficient

$$\sum_{h=2}^{k} \sum_{u=0}^{\min\{h-2, s\}} (h - u - 1) \binom{k}{h} \binom{s}{u} (-1)^{k-h+s-u} =: A(k, s)(-1)^{k+s}.$$

Now, we only need to prove that $A(k,s) = \binom{s+k-2}{s}$.

For notational convenience we will prove this statement for $k+1$. We start by splitting the first binomial coefficient, and then substituting $h \to h+1$ in the second part.

$$\begin{aligned}
A(k+1,s) &= \sum_{h=2}^{k+1}(-1)^h\left(\binom{k}{h}+\binom{k}{h-1}\right)\sum_{u=0}^{h-2}(-1)^u\binom{s}{u}(h-u-1)\\
&= A(k,s) + \sum_{h=1}^{k}(-1)^{h+1}\binom{k}{h}\sum_{u=0}^{h-1}(-1)^u\binom{s}{u}(h-u)\\
&= A(k,s) + \sum_{h=1}^{k}(-1)^{h+1}\binom{k}{h}\sum_{u=0}^{h-2}(-1)^u\binom{s}{u}(h-u)\\
&\qquad + \sum_{h=1}^{k}(-1)^{h+1}\binom{k}{h}(-1)^{h-1}\binom{s}{h-1}\\
&= \sum_{h=2}^{k}(-1)^{h+1}\binom{k}{h}\sum_{u=0}^{h-2}(-1)^u\binom{s}{u} + \sum_{h=1}^{k}\binom{k}{k-h}\binom{s}{h-1}.
\end{aligned}$$

Using Lemma 4.4 below on the first sum and the Vandermonde convolution $\sum_{i=0}^{r}\binom{x}{i}\binom{y}{r-i} = \binom{x+y}{r}$ on the last we get

$$A(k+1,s) = -\binom{s+k-1}{s+1} + \binom{s+k}{s+1} = \binom{s+k-1}{s}.$$

The formula (1.3) follows from the following lemma.

Lemma 4.4. *For any integers $k \geq 2$, $s \geq 0$ we have*

$$\sum_{h=2}^{k}(-1)^h\binom{k}{h}\sum_{u=0}^{h-2}(-1)^u\binom{s}{u} = \binom{s+k-1}{s+1}$$

Proof. First note that by standard binomial identities we have

$$\sum_{u=0}^{h-2}(-1)^u\binom{s}{u} = \sum_{u=0}^{h-2}\binom{-s+u-1}{u} = \binom{-s+h-2}{h-2} = (-1)^{h-2}\binom{s-1}{h-2}.$$

We may now use the Vandermonde convolution to get, with $j = h-2$,

$$\sum_{h=2}^{k}(-1)^h\binom{k}{h}\sum_{u=0}^{h-2}(-1)^u\binom{s}{u} = \sum_{j=0}^{k-2}\binom{k}{k-2-j}\binom{s-1}{j} = \binom{s+k-1}{k-2}.$$

□

This completes the proof of Theorem 1.3.

5. Proofs of corollaries

Proof of Corollary 2.1. When $m = 3$, the double sum in (1.3) is non-empty only if $i = 3$; in this case there is a single term with $s = 0$ and $k = 2$ and the result follows immediately.

Alternatively, we can use (1.2): $S_i^- - S_i^+ - 1 > 0$ is possible only with $S_i^- = 2$ and $S_i^+ = 0$; this requires that at least 2 states are larger than state i so $i = 3$, and in this case the probability is

$$\frac{1}{3}\,\mathbb{P}\big(S_3^- - S_3^+ = 2\big) = \frac{1}{3}\,\mathbb{P}\big(I_1^{(3)} = I_2^{(3)} = 1\big) = \frac{1}{3}(p_{(1)} - p_{(3)})(p_{(2)} - p_{(3)}). \qquad \square$$

Proof of Corollary 2.3. The fact that $q_{(m)} \geq q_{(m-1)} \geq \dots$ follows from (1.2) and a simple coupling argument. Furthermore, if $i \leq 2$, then $S_i^- \leq 1$ and $q_{(i)} = 0$.

For $i = m$, (1.3) simplifies to

$$\begin{aligned} mq_{(m)} &= \sum_{k=2}^{m-1} (-1)^k e_k\big((p_{(1)} - p_{(m)}), \dots, (p_{(m-1)} - p_{(m)})\big) \\ &= \prod_{j=1}^{m-1} \big(1 - (p_{(j)} - p_{(m)})\big) + \sum_{j=1}^{m-1} (p_{(j)} - p_{(m)}) - 1 \\ &= \prod_{j=1}^{m-1} \big(1 - (p_{(j)} - p_{(m)})\big) - mp_{(m)}, \end{aligned} \tag{5.1}$$

which is (2.1).

Furthermore, it follows from (2.1) (or (1.2)) that $q_{(m)}$ will increase if we decrease $p_{(m)}$ to 0 and simultaneously increase $p_{(1)}$, say. For $p_{(m)} = 0$, the product in (5.1) is largest when all p_j for $j < m$ are equal, i.e. $p_{(j)} = 1/(m-1)$ for $j < m$; in this case (5.1) yields $(1 - 1/(m-1))^{m-1}$. Hence, for any $p_{(m)} > 0$ and any i,

$$q_{(i)} \leq q_{(m)} < \frac{1}{m}\Big(1 - \frac{1}{m-1}\Big)^{m-1} < \frac{1}{m} e^{-1}. \tag{5.2}$$

Finally, for any i,

$$\mathbb{E}\big(S_i^- - S_i^+\big) = \sum_{p_j > p_i} (p_j - p_i) - \sum_{p_j < p_i} (p_i - p_j) = \sum_j (p_j - p_i) = 1 - mp_i. \tag{5.3}$$

Consequently, using (1.2),

$$\begin{aligned} mq_i = \mathbb{E}\big(S_i^- - S_i^+ - 1\big)_+ &\geq \mathbb{E}\left(\big(S_i^- - 1\big)_+ - S_i^+\right) \\ &= \mathbb{E}\big(S_i^- - 1 + \mathbf{1}\{S_i^- = 0\}\big) - \mathbb{E}\, S_i^+ \\ &= \mathbb{E}\, S_i^- - 1 + \mathbb{P}\big(S_i^- = 0\big) - \mathbb{E}\, S_i^+ \\ &= \mathbb{P}\big(S_i^- = 0\big) - mp_i = \prod_{p_j > p_i} \big(1 - (p_j - p_i)\big) - mp_i. \end{aligned} \tag{5.4}$$

If $x \in [0,1]$, then $1-x \ge e^{-x} - \frac{1}{2}x^2$, and it follows that

$$\prod_{p_j>p_i} \bigl(1-(p_j-p_i)\bigr) \ge \prod_{p_j>p_i} e^{-(p_j-p_i)} - \frac{1}{2}\sum_{p_j>p_i}(p_j-p_i)^2 \ge e^{-\sum_j p_j} - \frac{1}{2}\sum_j p_j^2,$$

which by (5.4) yields the lower bound in (2.2). □

Proof of Corollary 2.4. The expected number is $\sum_j q_j$, and (2.2) shows that $\sum_j q_j < \frac{m}{em} = e^{-1}$.

In the special case, (2.2) shows that $q_i \sim e^{-1}/m$ for each small state, and thus $\sum_j q_j \sim e^{-1}$. □

Proof of Corollary 2.5. It is well-known and easy to see that if $I \sim \mathrm{Be}(r)$ and $Y \sim \mathrm{Po}(r)$, then I and Y may be coupled with $\mathbb{E}\,|I-Y| = 2(e^{-r}-1+r) \le r^2$. Couple in this way $I_j^{(i)}$ with $Y_j \sim \mathrm{Po}(|p_i-p_j|)$, with the latter independent, and define $\widehat{S}^+ = \sum_{j:p_j<p_i} Y_j \sim \mathrm{Po}(\lambda^+)$ and $\widehat{S}^- = \sum_{j:p_j>p_i} Y_j \sim \mathrm{Po}(\lambda^-)$. Then, by (1.2), (2.3) and the triangle inequality,

$$\begin{aligned} m|q_i-\hat{q}_i| &\le \mathbb{E}\bigl|S_i^- - S_i^+ - (\widehat{S}^- - \widehat{S}^+)\bigr| \\ &\le \sum_{j:p_j>p_i} \mathbb{E}\,|I_j^{(i)} - Y_j| + \sum_{j:p_j<p_i} \mathbb{E}\,|I_j^{(i)} - Y_j| \\ &\le \sum_{j:p_j>p_i} (p_j-p_i)^2 + \sum_{j:p_j<p_i} (p_i-p_j)^2. \end{aligned}$$

This proves (2.3), which immediately yields (2.4). To obtain (2.5) we observe that the sum is $\mathbb{E}\bigl(\widehat{S}^+ + 1 - \widehat{S}^-\bigr)_+ = -\mathbb{E}\bigl(\widehat{S}^- - \widehat{S}^+ - 1\bigr)_-$, and thus the difference between the sums in (2.4) and (2.5) equals

$$\begin{aligned} &\mathbb{E}\bigl(\widehat{S}^- - \widehat{S}^+ - 1\bigr)_+ + \mathbb{E}\bigl(\widehat{S}^- - \widehat{S}^+ - 1\bigr)_- = \mathbb{E}\bigl(\widehat{S}^- - \widehat{S}^+ - 1\bigr) \\ &= \sum_{j:p_j>p_i}(p_j-p_i) - \sum_{j:p_j<p_i}(p_i-p_j) - 1 = \sum_{j=1}^m (p_j-p_i) - 1 \\ &= 1 - mp_i - 1 = -mp_i. \end{aligned}$$

□

6. Proofs of results on average probabilities

We let $\Delta_a^m := \{(x_1,\dots,x_m) \in [0,\infty)^m : \sum_i x_i = a\}$ and $\Delta_{\le a}^m := \{(x_1,\dots,x_m) \in [0,\infty)^m : \sum_i x_i \le a\}$. When integrating over Δ_a^m, we use the measure $\mathrm{d}\mathbf{x} := \mathrm{d}x_1 \cdots \mathrm{d}x_{m-1}$; this is thus the same as integrating over $\Delta_{\le a}^{m-1}$ with Lebesgue measure, keeping $x_m = 1 - \sum_1^{m-1} x_i$. Note that the volume of Δ_a^m equals the volume of $\Delta_{\le a}^{m-1}$, i.e. $a^{m-1}/(m-1)!$. Hence, the uniform probability measure on $\Delta^m = \Delta_1^m$ is $(m-1)!\,\mathrm{d}\mathbf{x}$.

More generally, we have the well-known Dirichlet integral

$$\int_{\Delta_a^m} x_1^{\alpha_1-1} \cdots x_m^{\alpha_m-1}\,\mathrm{d}\mathbf{x} = a^{\alpha_1+\dots+\alpha_m-1} \frac{\prod_{i=1}^m \Gamma(\alpha_i)}{\Gamma\bigl(\sum_{i=1}^m \alpha_i\bigr)} \tag{6.1}$$

for any $\alpha_1, \dots, \alpha_m > 0$. (For $m = 2$ this is the standard Beta integral and the general case follows easily by induction. An alternative, probabilistic, standard proof is to let $T_1, \dots, T_m$ be independent Exp(1) variables and evaluate $\mathbb{E}(T_1^{\alpha_1-1} \cdots T_m^{\alpha_m-1})$ by conditioning on $T_1 + \cdots + T_m$.)

Proof of Theorem 3.1. Recall from Corollary 2.3 that if we assume $p_m \le p_i$ for all i and let $r_i := p_i - p_m$, then $q_m = \frac{1}{m}\prod_{i=1}^{m-1}(1-r_i) - p_m$. Choosing a vector $(p_1, \dots, p_m)$ uniformly from Δ_1^m, there are m possibilities for the position of the minimum coordinate; by symmetry it suffices to consider the case when p_m is the minimum (multiplying below by a factor m). Then the vector $(r_1, \dots, r_{m-1})$ is uniformly distributed in $\Delta_{1-mp_m}^{m-1}$ and $\mathrm{d}\mathbf{x} = \mathrm{d}p_m \,\mathrm{d}\mathbf{r}$. We thus get

$$\mathbb{E}\, q_{(m)} = m(m-1)! \int_0^{1/m} \int_{\mathbf{r}\in\Delta_{1-mp_m}^{m-1}} \left(\frac{1}{m} \prod_{i=1}^{m-1} (1-r_i) - p_m \right) \mathrm{d}\mathbf{r}\ \mathrm{d}p_m. \tag{6.2}$$

We treat the two terms in the bracket separately. First, using symmetry and (6.1),

$$\begin{aligned}
(m-1)! &\int_0^{1/m} \int_{\mathbf{r}\in\Delta_{1-mp_m}^{m-1}} \prod_{i=1}^{m-1} (1-r_i)\, \mathrm{d}\mathbf{r}\ \mathrm{d}p_m \\
&= (m-1)! \sum_{k=0}^{m-1} \binom{m-1}{k} (-1)^k \int_0^{1/m} \int_{\mathbf{r}\in\Delta_{1-mp_m}^{m-1}} \prod_{i=1}^{k} r_i \,\mathrm{d}\mathbf{r}\ \mathrm{d}p_m \\
&= \sum_{k=0}^{m-1} \binom{m-1}{k} (-1)^k \int_0^{1/m} (1-mp_m)^{k+m-2} \frac{(m-1)!}{\Gamma(k+m-1)}\, \mathrm{d}p_m \\
&= \sum_{k=0}^{m-1} \binom{m-1}{k} (-1)^k \frac{(m-1)!}{m\Gamma(k+m)} \\
&= \frac{1}{m} \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} \frac{1}{m^{\overline{k}}}.
\end{aligned} \tag{6.3}$$

Similarly, the second term becomes

$$\begin{aligned}
m(m-1)! \int_0^{1/m} \int_{\mathbf{r}\in\Delta_{1-mp_m}^{m-1}} p_m \,\mathrm{d}\mathbf{r}\ \mathrm{d}p_m &= m(m-1) \int_0^{1/m} (1-mp_m)^{m-2} p_m \,\mathrm{d}p_m \\
&= \frac{m-1}{m} \int_0^1 (1-x)^{m-2} x \,\mathrm{d}x = \frac{1}{m^2}.
\end{aligned} \tag{6.4}$$

The formula (3.1) follows from (6.3) and (6.4), noting that the first two terms in the first sum equal $1 - \frac{m-1}{m} = \frac{1}{m}$, which cancels the term $-1/m^2$.

The asymptotic expansion (3.2) is easy to deduce from (3.1). (One can also easily obtain further terms.)

Since $q_{(m)} \le e^{-1}/m$ a.s., we have $\mathbb{E}\,|e^{-1} - mq_{(m)}| = \mathbb{E}(e^{-1} - mq_{(m)}) = e^{-1} - m\,\mathbb{E}\,q_{(m)} \to 0$, and thus $mq_{(m)} \xrightarrow{\text{p}} e^{-1}$. $\square$

Proof of Theorem 3.3. Take $i = m$. We take expectations in (1.4) and obtain by symmetry

$$\mathbb{E}\,q_m = \frac{1}{m}\sum_{s=0}^{m-3}\sum_{k=2}^{m-s-1}(-1)^{s+k}\binom{s+k-2}{s}\binom{m-1}{s,k,m-1-s-k} \cdot \mathbb{E}\left(\prod_{j=1}^{s}(p_m-p_j)_+\prod_{l=1}^{k}(p_{s+l}-p_m)_+\right). \tag{6.5}$$

We use the standard method of generating uniform $p_1,\dots,p_m$ on Δ^m by taking i.i.d. exponential random variables $T_1,\dots,T_m \sim \mathrm{Exp}(1)$ and letting $p_i := T_i/S_m$ with $S_m := \sum_{i=1}^m T_i$. Recall that then $(p_1,\dots,p_m)$ and S_m are independent, and that S_m has a Gamma$(m,1)$ distribution with $\mathbb{E}\,S_m^\alpha = \Gamma(m+\alpha)/\Gamma(m)$ for $\alpha \in \mathbb{N}$. Hence, by conditioning on $T = T_m$ and using independence,

$$\begin{aligned}
\mathbb{E}&\left(\prod_{j=1}^{s}(p_m-p_j)_+\prod_{l=1}^{k}(p_{s+l}-p_m)_+\right)\\
&= \frac{\mathbb{E}\left(\prod_{j=1}^{s}(T_m-T_j)_+\prod_{l=1}^{k}(T_{s+l}-T_m)_+\right)}{\mathbb{E}\,S_m^{s+k}}\\
&= \frac{\Gamma(m)}{\Gamma(m+k+s)}\,\mathbb{E}\Bigl(\bigl(\mathbb{E}((T-T_1)_+ \mid T)\bigr)^s\bigl(\mathbb{E}((T_2-T)_+ \mid T)\bigr)^k\Bigr). \qquad (6.6)
\end{aligned}$$

For any $t > 0$ and $j \ge 1$ we have

$$\mathbb{E}(t-T_j)_+ = \int_0^t (t-x)e^{-x}\,\mathrm{d}x = e^{-t} - 1 + t, \tag{6.7}$$

$$\mathbb{E}(T_j-t)_+ = \int_t^\infty (x-t)e^{-x}\,\mathrm{d}x = e^{-t}. \tag{6.8}$$

Hence, the final expectation in (6.6) equals

$$\begin{aligned}
\mathbb{E}\Bigl(\bigl(e^{-T}-1+T\bigr)^s e^{-kT}\Bigr) &= \int_0^\infty \bigl(e^{-t}-1+t\bigr)^s e^{-(k+1)t}\,\mathrm{d}t\\
&= \int_0^\infty \sum_{i+j\le s}\binom{s}{i,j,s-i-j}e^{-it}t^j(-1)^{s-i-j}e^{-(k+1)t}\,\mathrm{d}t\\
&= \sum_{i+j\le s}(-1)^{s-i-j}\binom{s}{i,j,s-i-j}\int_0^\infty t^j e^{-(i+k+1)t}\,\mathrm{d}t\\
&= \sum_{i+j\le s}(-1)^{s-i-j}\binom{s}{i,j,s-i-j}\frac{j!}{(i+k+1)^{j+1}}. \qquad (6.9)
\end{aligned}$$

The result now follows from (6.5)–(6.9). □

Proof of Theorem 3.5. Note first that the expectation in (3.4) can be evaluated as in (3.5)–(3.6). For (3.5) this is immediate; for (3.6) it follows from

$$\mathbb{E}\bigl(\widehat{S}^- - \widehat{S}^+ - 1\bigr)_+ - \mathbb{E}\bigl(\widehat{S}^- - \widehat{S}^+ - 1\bigr)_- = \mathbb{E}\bigl(\widehat{S}^- - \widehat{S}^+ - 1\bigr) = \lambda^- - \lambda^+ - 1.$$

We let, as in the proof of Theorem 3.3, $p_i := T_i/S_m$ with $T_i \sim \mathrm{Exp}(1)$ i.i.d. and $S_m := \sum_{i=1}^m T_i$. Consider state 1 and condition on T_1, leaving $T_2, T_3, \dots$ i.i.d. Exp(1). We define $\lambda^\pm := \sum_j (p_1 - p_j)_\pm$ as in Corollary 2.5 (with $i = 1$). As $m \to \infty$, the law of large numbers shows that a.s., using (6.7)–(6.8),

$$\frac{S_m}{m} = \frac{T_1}{m} + \sum_{j=2}^m \frac{T_j}{m} \to 0 + \mathbb{E}\, T_2 = 1, \tag{6.10}$$

$$\begin{aligned} \lambda^+ &= \sum_{j=2}^m \frac{(T_1 - T_j)_+}{S_m} = \frac{m}{S_m}\sum_{j=2}^m \frac{(T_1 - T_j)_+}{m} \\ &\to \mathbb{E}\bigl((T_1 - T_2)_+ \mid T_1\bigr) = e^{-T_1} - 1 + T_1, \end{aligned} \tag{6.11}$$

$$\begin{aligned} \lambda^- &= \sum_{j=2}^m \frac{(T_1 - T_j)_-}{S_m} = \frac{m}{S_m}\sum_{j=2}^m \frac{(T_1 - T_j)_-}{m} \\ &\to \mathbb{E}\bigl((T_1 - T_2)_- \mid T_1\bigr) = e^{-T_1}, \end{aligned} \tag{6.12}$$

$$\sum_{j=1}^m (p_j - p_1)^2 = \frac{m}{S_m^2}\sum_{j=2}^m \frac{(T_j - T_1)^2}{m} = O\Bigl(\frac{1}{m}\Bigr). \tag{6.13}$$

First, by (6.10),

$$mp_1 = \frac{m}{S_m} T_1 \to T_1. \tag{6.14}$$

Next, we apply Corollary 2.5, noting that $\hat{q}_1 = \Phi(\lambda^-, \lambda^+)/m$. Hence, by (2.6) and (6.13), $mq_1 - \Phi(\lambda^-, \lambda^+) \to 0$ a.s. For any $\lambda_1, \lambda_2 > 0$, we can couple $\widehat{S}_1 \sim \mathrm{Po}(\lambda_1)$ and $\widehat{S}_2 \sim \mathrm{Po}(\lambda_2)$ such that $\mathbb{E}\,|\widehat{S}_1 - \widehat{S}_2| \le |\lambda_1 - \lambda_2|$, and it follows from (3.4) and the triangle inequality that $|\Phi(\lambda_1^-, \lambda_1^+) - \Phi(\lambda_2^-, \lambda_2^+)| \le |\lambda_1^- - \lambda_2^-| + |\lambda_1^+ - \lambda_2^+|$ for any $\lambda_1^-, \lambda_1^+, \lambda_2^-, \lambda_2^+$. Hence (6.11)–(6.12) imply that $\Phi(\lambda^-, \lambda^+) \to \Psi(T_1)$ a.s. Consequently, a.s.

$$mq_1 \to \Psi(T_1). \tag{6.15}$$

The limit (3.7) follows from (6.14)–(6.15). Since $mq_1 \le e^{-1}$ a.s. by Corollary 2.3, (3.8) follows by dominated convergence toghether with (3.5). □

7. Further comments

7.1. **Several states at once.** Several states may suffer from the Alabama paradox at the same time. A simple example is given in Figure 3.

This can be analysed in the same way using the methods in Section 4. The typical case is when two states i and j are in positions 0 and 1, and both

state	pop.	μ_i	seats
A	28	1.40	1
B	27	1.35	1
C	27	1.35	1
D	9	0.45	**1**
E	9	0.45	**1**
sum	100	5.00	5

5 seats

state	pop.	μ_i	seats
A	28	1.68	**2**
B	27	1.62	**2**
C	27	1.62	**2**
D	9	0.54	0
E	9	0.54	0
sum	100	6.00	6

6 seats

FIGURE 3. A double Alabama paradox for states D and E. Numbers in boldface are rounded up.

are overtaken by three other runners. However, other similar configurations are possible, and we leave the details to the reader except for two simple examples.

Example 7.1. Suppose that there are 5 states, with $p_1 \geq p_2 \geq p_3 \geq p_4 \geq p_5$ and (p_i) linearly independent over $\mathbb{Q}$. If a double Alabama paradox occurs, it has to be for states 4 and 5, with one of them in position 0, the other in position 1, and the three others overtaking both of them. Letting x be the distance between the runners 4 and 5, we find the probability, using uniform distribution as before,

$$\begin{aligned}&\frac{1}{5}\int_0^{p_3-p_4}(p_3-p_4-x)(p_2-p_4-x)(p_1-p_4-x)\,\mathrm{d}x\\&+\frac{1}{5}\int_0^{p_4-p_5}(p_3-p_4)(p_2-p_4)(p_1-p_4)\,\mathrm{d}x\\&+\frac{1}{5}\int_{p_4-p_5}^{p_3-p_5}(p_3-p_5-x)(p_2-p_5-x)(p_1-p_5-x)\,\mathrm{d}x.\end{aligned}$$

The integrals are easily evaluated (preferably by computer), but the resulting polynomial in $p_1, \dots, p_5$ does not look particularly nice or illuminating so we omit it. In the extreme case $p_1 \approx p_2 \approx p_3 \approx 1/3$, $p_4 \approx p_5 \approx 0$, the probability becomes 1/810.

Example 7.2. Consider the extreme case in Corollary 2.4, with $m - o(m)$ small states and $o(m)$ medium-size states with the bulk of the population. In this case, the flag is most likely carried by a small state, say i. The probability that it is overtaken by a large state j is $p_j - p_i \approx p_j$, and we can approximate the distributions of the number M of states overtaking it by a Poisson distribution with mean $\sum_j (p_j - p_i)_+ \approx \sum_j p_j = 1$. Thus i loses a seat if $M > 1$. Furthermore, since there are very many small states, their runners are very narrowly spaced, and most likely the runners in positions $1, \dots, M-2$ are also small states, and are also overtaken by the same M states as i; in this case they all lose a seat.

state	pop.	1	2	3	4	5	6	7
A	6	**0.6**	1.2	**1.8**	*2.4*	3.0	*3.6*	4.2
B	3	0.3	**0.6**	**0.9**	1.2	*1.5*	**1.8**	2.1
C	1	0.1	0.2	0.3	*0.4*	*0.5*	*0.6*	**0.7**

FIGURE 4. Ties. Numbers in boldface are rounded up; numbers in italics may be rounded up or down, as determined by lot.

Consequently, if the random variable X_m is the number of states suffering from the Alabama paradox when n is increased (from a random value), then as $m \to \infty$ (in this case), $X_m \xrightarrow{\mathrm{d}} X := (Y-1)_+$ with $Y \sim \mathrm{Po}(1)$. Note that $\mathbb{P}(X > 0) = \mathbb{P}(Y > 1) = 1 - 2e^{-1}$ and $\mathbb{E}\, X = \mathbb{E}(Y - 1 + \mathbf{1}\{Y = 0\})$ $= \mathbb{P}(Y = 0) = e^{-1}$.

7.2. **The linearly dependent case.** The proof above uses Weyl's theorem and thus requires that $p_1, \dots, p_m$ are linearly independent over $\mathbb{Q}$; indeed, as said in Section 1, simple examples shows that Theorem 1.3 does not hold for arbitrary p_i. We note also that some of the corollaries might be far from true in the case of rational relative sizes, see Proposition 7.4 below. Nevertheless, it is possible to use much of the argument above also for the linearly dependent case. We sketch this below, leaving many details to the reader.

Note first that in general it may happen that the remainders $\mu_i - \lfloor \mu_i \rfloor$ happen to be equal for two or more states, and it may be necessary to round up one or several of these and round down the others; in this case, the choice is determined by lot. A simple example is given in Figure 4; note that state C may suffer the Alabama paradox either when increasing from 4 to 5 seats or from 5 to 6, but not at both times; the probability is 1/4 each time. For the asymptotic analysis this complication is no real problem, however, since it suffices to consider the expectation $\mathbb{E}\, \nu_i(N)$, i.e., the sum over $n \le N$ of the probability of an Alabama paradox at time n. Indeed, although the example in Figure 4 shows that there may be a dependency between the occurrence of the Alabama paradox for some number n of seats and the next number $n+1$, more distant occurrences are independent; i.e., the random sequence of occurrences of the paradox is 1-dependent. Hence the variance of the total number is $O(N)$, and by considering odd and even n separately, which yields two subsequences of independent random indicators, it is easily seen that $(\nu_i(N) - \mathbb{E}\, \nu_i(N))/N \to 0$ almost surely.

Remark 7.3. If two or several states have exactly the same population, it may be necessary to draw lots between them for both n seats and $n+1$. If we do this independently, then some state might lose a seat by being unlucky the second time. This obvious consequence of drawing lots is not an example of the Alabama paradox and should be disregarded. (For example, we may list the states with the same population in some random order, once and for all, and use this as a priority list for each n.)

Let $\mathbf{p} := (p_1, \dots, p_m)$. We now use q_i for the true probability and $\tilde{q}_i$ for the value in (1.2) for any $p_1, \dots, p_m$. Consider again, for notational convenience, state 1. The argument in Section 4 shows that, with L_n given by (4.6) (noting that the actual position may be different when there are ties), for some functions f, g and h and with $z_j = (p_j - p_1)/m$, for simplicity assuming that N is a multiple of m,

$$\begin{aligned}
\mathbb{E}\,\nu_1(N) &= \sum_{n=0}^{N-1} f\big(\{n(p_2-p_1)\}, \dots, \{n(p_m-p_1)\}, L_n\big) \\
&= \sum_{n=0}^{N-1} g\big(\{nz_2\}, \dots, \{nz_m\}, \mathrm{Mod}_m(n)\big). \\
&= \sum_{k=1}^{N/m} mh\big(\{k(p_2-p_1)\}, \dots, \{k(p_m-p_1)\}\big). \qquad (7.1)
\end{aligned}$$

(To see the last equality, write $n = (k-1)m + l$ and define $h(x_2, \dots x_m) := \sum_{l=0}^{m-1} g(\{x_2 + (l-m)z_2\}, \dots, \{x_m + (l-m)z_m\}, l)/m$.) Lemma 4.2 does not apply when $p_1, \dots, p_m$ are linearly dependent over $\mathbb{Q}$, but the proof of it sketched above shows that the sequence $\big(\{n(p_2-p_1)\}, \dots, \{n(p_m - p_1)\}\big)$ is uniformly distributed on a subgroup of $[0,1)^{m-1}$; more precisely, the empirical distributions converge to the uniform probability measure $\mu_{\mathbf{p}}$ on this subgroup, which has Fourier coefficients given by

$$\widehat{\mu}_{\mathbf{p}}(a_2, \dots, a_m) = \begin{cases} 1 & \text{if } \sum_{j=2}^m a_j(p_j - p_1) \in \mathbb{Z}, \\ 0 & \text{otherwise.} \end{cases} \qquad (7.2)$$

The functions f, g and h are linear combinations of products of indicators and are a.e. continuous; moreover, h is $\mu_{\mathbf{p}}$-a.e. continuous. Hence, (7.1) shows that $\mathbb{E}\,\nu_1(N)/N \to \int h \,\mathrm{d}\mu_{\mathbf{p}}$, showing the existence of the limit $q_1 = \int h \,\mathrm{d}\mu_{\mathbf{p}}$ in general. Note that in this notation, the value $\tilde{q}_1$ in (1.2) is $\tilde{q}_1 = \int h \,\mathrm{d}\mu$, where μ is the uniform distribution on $[0,1)^{m-1}$.

The functions f, g and h depend on $\mathbf{p}$, cf. (4.2)–(4.3), but if we write $h_{\mathbf{p}}$, then $h_{\mathbf{p}_k} \to h_{\mathbf{p}}$ a.e. for any sequence of population distributions (for a fixed number of states) $\mathbf{p}_1, \mathbf{p}_2, \dots$, with $\mathbf{p}_k \to \mathbf{p}$. Let $\mathbf{p}_k = (p_{1k}, \dots, p_{mk})$. If further, for every integer vector $(a_1, \dots, a_m) \neq 0$, we have $\sum_{i=1}^m a_i p_{ik} \neq 0$ for all large k, then it follows from (7.2) that $\widehat{\mu}_{\mathbf{p}_k}(a_2, \dots, a_m) \to 0$ for all $(a_2, \dots, a_m) \neq 0$, and thus $\mu_{\mathbf{p}_k} \to \mu = \mu_{\mathbf{p}}$. It follows, e.g. by [2, Theorem 5.5] that $q_{1,k} = \int h_{\mathbf{p}_k} \,\mathrm{d}\mu_{\mathbf{p}_k} \to \int h_{\mathbf{p}} \,\mathrm{d}\mu_{\mathbf{p}}$ and also $\int h_{\mathbf{p}_k} \,\mathrm{d}\mu \to \int h_{\mathbf{p}} \,\mathrm{d}\mu_{\mathbf{p}}$ so $q_{1,k} - \int h_{\mathbf{p}_k} \,\mathrm{d}\mu \to 0$. The claim in Remark 1.2 now follows, since otherwise one could, for some $\varepsilon > 0$, find such a sequence $\mathbf{p}_k$ of distributions with $|q_{1,k} - \int h_{\mathbf{p}_k} \,\mathrm{d}\mu| \geq \varepsilon$ and (by taking a subsequence) $\mathbf{p}_k \to \mathbf{p}$ for some $\mathbf{p}$; a contradiction.

We end with a couple of counterexamples in the rational case. The upper bound $q_i < \frac{1}{m}e^{-1}$ in Corollary 2.3 is not true in general, not even for $m = 3$. As an easy example one may study three states with $p_1 = p_2 = 3/7$ and $p_3 = 1/7$, where $q_3 = 1/7 > \frac{1}{3e}$, since the smallest state suffers when n increases from 3 to 4 (mod 7).

Moreover, the upper bound $1/e$ on the expected number of states suffering from the paradox in Corollary 2.4 is not true in general. Consider the case when $p_1 = \cdots = p_6 = 1/7$ and $p_7 = \cdots = p_{106} = 1/700$. When e.g. the number of seats changes from 42 to 43 all the 6 large states will change from 6 to 7 seats. Thus 5 of the small states will suffer from the Alabama paradox. The paradox will happen 98 times during the period of length 700; 90 of these 5 small states will suffer, but the number is smaller close to the beginning and end of the period ($n = 7$, 14, 21, 28 or 671, 678, 685, 692 (mod 700)) and the expected number of states to suffer from the Alabama paradox for a random number of seats is

$$\frac{90 \cdot 5 + 2 \cdot (1 + 2 + 3 + 4)}{700} = \frac{47}{70},$$

which is much larger than $1/e$. This can be generalized in the following way.

Proposition 7.4. *Let $x \geq 2$ and y be positive integers, with x relatively prime to $y - 1$ and $x^2 - 3x < y$. Assume that the number of states is $m = y + x - 1$ with relative sizes $p_1 = \cdots = p_{x-1} = \frac{1}{x}$ and $p_x = \cdots = p_m = \frac{1}{x \cdot y}$. Then the expected number of states suffering from the Alabama paradox equals*

$$\frac{(x-2)(y-x+1)}{xy}, \tag{7.3}$$

which in particular can be made arbitrarily close to 1.

Proof. The values of n (mod xy) where the Alabama paradox might occur are, e.g. by considering runners as in Section 4, $ax + b$, for integers $1 \leq a \leq y - 2$ and $0 \leq b \leq x - 1$ such that $ax + b > by$ and $ax + b + 1 < (b + 1)y$, i.e. $b = \lfloor ax/(y - 1) \rfloor$. There are $y - 2$ such values $ax + b$, one for each $a \in \{1, \dots, y - 2\}$.

If $x - 2 \leq a \leq y - x + 1$, then for $n = ax + b$, with b as above, the $x - 1$ large states will get a seats each and the remaining $a + b$ seats will go to $a + b$ of the small states. For $n = ax + b + 1$, the larger states will get $a + 1$ seats each and only $a + b - x + 2$ seats are left to the small states; thus $x - 2$ small states will suffer from the Alabama paradox. If $1 \leq a \leq x - 3$, then $b = 0$, and the number of small states receiving a seat will drop from $a + b = a$ to 0 as n increases from $ax + b$ to $ax + b + 1$. Finally, if $y - x + 2 \leq a \leq y - 2$, then $b = x - 1$ and the number will drop from y to $a + b - x + 2 = a + 1$, so $y - 1 - a$ states will suffer. Summing these numbers give a total of $(y - 2x + 4)(x - 2) + 2\sum_{i=1}^{x-3} i = (y - x + 1)(x - 2)$ states suffering in the period xy.

Taking e.g. $y = x^2$ and letting $x \to \infty$, the expected number (7.3) tends to 1. $\square$

We thank Warren D. Smith for asking a question that made us produce Proposition 7.4. It would be interesting to see a proof of 1 being a general upper bound or an example to the contrary.

References

[1] M. L. Balinski & H. P. Young, *Fair Representation.* 2nd ed., Brookings Institution Press, Washington, D.C., 2001.

[2] P. Billingsley, *Convergence of Probability Measures.* Wiley, New York, 1968.

[3] M. Fehndrich, Paradoxien von Hare/Niemeyer. Wahlrecht.de. `http://www.wahlrecht.de/verfahren/paradoxien/index.html`

[4] L. Grafakos, *Classical and Modern Fourier Analysis.* Pearson, Upper Saddle River, NJ, 2004.

[5] S. Janson, Asymptotic bias of some election methods. Preprint, 2011. `arXiv:1110.6369.`

Department of Mathematics, Uppsala University, P.O. Box 480, SE-751 06, Uppsala, Sweden.

E-mail address: `svante@math.uu.se`

URL: `http://www.math.uu.se/~svante/`

Department of Mathematics, KTH – Royal Institute of Technology, SE-100 44, Stockholm, Sweden.

E-mail address: `linusson@math.kth.se`

URL: `http://www.math.kth.se/~linusson/`