Optimal Congressional Apportionment

Robert A. Agnew

1. INTRODUCTION. After two centuries of experimentation with alternative allocation rules, and considerable political maneuvering, the issue of the "best" method for allocating seats in the U.S. House of Representatives remains unsettled, although the field has narrowed considerably to so-called "divisor" methods. Mathematically, we have s states, house size h, state populations $p_1, \dots, p_s$, total population $p = \sum_{i=1}^{s} p_i$, "quota" $q_i = hp_i/p$ for state i, and integer allocation of seats a_i for state i; currently, of course, $s = 50$ and $h = 435$. In addition, we have constitutionally mandated lower bounds $a_i \geq 1$ (i.e., each state shall have at least one seat), and there are currently inactive upper bounds $a_i \leq p_i/30{,}000$ (i.e., the number of representatives in each state shall not exceed one per thirty-thousand). Beyond these constraints, Article I of the U.S. Constitution simply says that seats "shall be apportioned among the several states according to their respective numbers." In other words, a_i should be "close to" q_i.

To avoid controversy, the allocation methodology should be perceived as reasonably balanced between large and small states. It should also be sensible in terms of avoiding paradoxes associated with addition of new states, increase in house size, or differential state growth rates; divisor methods are generally free of these impediments [**1**]. The Hill method, in use by law since 1941, is a divisor method. The Webster method, in use during some previous decades, is a divisor method. Another divisor method due to Jefferson was used early on but was abandoned because it clearly favored large states. A non-divisor method due to Hamilton was used in the 1800s, but it was abandoned when it gave rise to nonsensical paradoxes.

The purpose of this article is to expand the list of divisor methods to include those associated with the logarithmic and identric means. We show that these two methods stem from mirror-image objective functions in terms of optimization, in the same sense that the Hill and Webster methods have mirror-image objective functions. We then explore the connections of the various objective functions to information theory and statistics, concluding that the identric mean and arithmetic mean (Webster) objective functions are the most natural, and moreover that the former has certain theoretical advantages. Finally, we compare optimal congressional allocations associated with the four means: geometric, logarithmic, identric, and arithmetic.

Our principal reference on congressional apportionment is the comprehensive monograph by Balinski and Young [**1**], which provides rich historical context, empirical support for the Webster method, and Appendix A which lays out the mathematical structure, including optimization criteria. We adopt their notation and avoidance of positive lower bounds, which are easily added to divisor calculations; we also ignore the upper bounds which are currently inactive and which could easily be added to the calculations if necessary. Two recent articles in this MONTHLY have dealt with the apportionment problem. Grimmett [**4**] proposed a randomized (lottery) scheme which guarantees quota in an expected value sense in the absence of lower bounds. Balinski [**2**] presented a general theory of coherence for apportionment problems, including the congressional problem.

2. DIVISOR METHODS. We assume that d is an increasing function on the non-negative integers such that $d(a) \in [a, a+1]$ for every a. Different d-functions lead to alternative allocation methods. We next introduce a "divisor," which is a notional ratio of population heads per congressional seat. For real divisor $x > 0$, we put $a_i = \lfloor p_i/x \rfloor$ if $p_i/x < d(\lfloor p_i/x \rfloor)$ and $a_i = \lfloor p_i/x \rfloor + 1$ if $p_i/x > d(\lfloor p_i/x \rfloor)$. If $p_i/x = d(\lfloor p_i/x \rfloor)$ we have a tie where a_i can be either $\lfloor p_i/x \rfloor$ or $\lfloor p_i/x \rfloor + 1$. Equivalently, we choose a_i such that $d(a_i - 1) \le p_i/x \le d(a_i)$, or $p_i/d(a_i - 1) \ge x \ge p_i/d(a_i)$, or $q_i/d(a_i - 1) \ge xh/p \ge q_i/d(a_i)$, where the terms including $d(a_i - 1)$ only apply when $a_i \ge 1$. Another characterization of a divisor solution is that $q_i/d(a_i - 1) \ge q_j/d(a_j)$ whenever $a_i \ge 1$ and $i \ne j$; this relationship will be important in the next section.

Our objective is to find a constrained divisor solution where $\sum a_i = h$ by adjusting the divisor x. If the state populations are distinct, as they always have been, there will normally be one such solution $a_1, \dots, a_s$, although the associated divisor is not unique. Algorithmically, we have to deal with ties in a systematic way. Our approach is to put $a_i = \lfloor p_i/x \rfloor$ if $p_i/x \le d(\lfloor p_i/x \rfloor)$, $a_i = \lfloor p_i/x \rfloor + 1$ otherwise, and, starting with a sufficiently large divisor to ensure $\sum a_i < h$, to approximate the minimum divisor satisfying the constraint $\sum a_i = h$ using *Microsoft Excel® Standard Solver*. This algorithm is straightforward, but there are theoretical instances where it can fail to deliver a proper constrained solution, even when the state populations are distinct.

For example, suppose we have two states, i and j, with $p_i/x = d(\lfloor p_i/x \rfloor)$, $p_j/x = d\left(\lfloor p_j/x \rfloor\right)$, and $\sum a_k = h - 1$. If x is decreased slightly, the algorithm will increase both a_i and a_j by one, yielding $\sum a_k = h + 1$, i.e., we will have skipped over our target house size h. The problem is that when $p_i/x = d(\lfloor p_i/x \rfloor)$, a_i can actually be either $\lfloor p_i/x \rfloor$ or $\lfloor p_i/x \rfloor + 1$; in the instance posed, either a_i or a_j can be increased to meet the house size constraint, so we have a nonunique solution, which is of course politically undesirable. Fortunately, this situation appears to be very unlikely, and in fact it did not occur in any of our algorithmic calculations.

Attention has naturally focused on d-functions derived from means between a and $a + 1$ (see Table A.3 in [**1**]). The following table adds two means to the usual list.

Table 1. Comparison of divisor methods

Method	Mean	$d(a)$
Adams	Minimum	a
Dean	Harmonic	$a(a+1)/(a+1/2)$
Hill	Geometric	$\sqrt{a(a+1)}$
	Logarithmic	$1/\ln\big((a+1)/a\big)$
	Identric	$(a+1)^{a+1}/(ea^a)$
Webster	Arithmetic	$a + 1/2$
Jefferson	Maximum	$a + 1$

Recall that for any a and b satisfying $0 < a < b$, we have

$$\sqrt{ab} \le \frac{b-a}{\ln b - \ln a} \le \frac{1}{e}\left(\frac{b^b}{a^a}\right)^{1/(b-a)} \le \frac{a+b}{2},$$

so the logarithmic and identric means fit into the natural hierarchy and the table contains the cases where $b = a + 1$. We next show that these two means correspond to mirror-image, information-theoretic objective functions in an optimization context. Note that $d(0) = 0$ for every entry in Table 1 up through logarithmic, so the constitutional lower bounds are automatically satisfied; for the last three methods $d(0) > 0$ (for the identric mean, $d(0) = e^{-1}$), so lower bounds must be imposed on these algorithms.

3. OPTIMIZATION. It is demonstrated in [**1**] that the arithmetic mean (Webster) method results from minimizing

$$\sum p_i \left(\frac{a_i}{p_i} - \frac{h}{p}\right)^2 = \left(\frac{h}{p}\right) \sum \frac{(a_i - q_i)^2}{q_i},$$

subject to $\sum a_i = h$ and $a_i \geq 0$. It is also indicated there that the geometric mean (Hill) method results from minimizing

$$\sum a_i \left(\frac{p_i}{a_i} - \frac{p}{h}\right)^2 = \left(\frac{p}{h}\right)^2 \sum \frac{(q_i - a_i)^2}{a_i}.$$

Fundamentally, then, these two objective functions are mirror images in terms of the roles played by the a_i and q_i. We now provide a more comprehensive result.

Proposition 1. *The following four divisor methods result from minimizing the indicated objective functions, subject to* $\sum a_i = h$ *and* $a_i \geq 0$.

Geometric Mean	$\sum (q_i - a_i)^2/a_i$	(1)
Logarithmic Mean	$\sum q_i(\ln q_i - \ln a_i)$	(2)
Identric Mean	$\sum a_i(\ln a_i - \ln q_i)$	(3)
Arithmetic Mean	$\sum (a_i - q_i)^2/q_i$	(4)

Proof. For any minimizing objective function $\sum f_i(a_i)$, optimality requires

$$f_i(a_i) + f_j(a_j) \leq f_i(a_i - 1) + f_j(a_j + 1),$$

or

$$f_i(a_i - 1) - f_i(a_i) \geq f_j(a_j) - f_j(a_j + 1),$$

whenever $a_i \geq 1$ and $i \neq j$. Otherwise, we can reduce the objective function by switching one seat from state i to state j. For objective function $\sum (a_i - q_i)^2/q_i$, we have

$$f_i(a_i - 1) - f_i(a_i) = 2 - \frac{2(a_i - 1/2)}{q_i}$$

and

$$f_j(a_j) - f_j(a_j + 1) = 2 - \frac{2(a_j + 1/2)}{q_j},$$

so we must have

$$-\frac{2(a_i - 1/2)}{q_i} \geq -\frac{2(a_j + 1/2)}{q_j},$$

or

$$\frac{q_i}{d(a_i - 1)} \geq \frac{q_j}{d(a_j)}$$

with $d(a) = a + 1/2$, whenever $a_i \geq 1$ and $i \neq j$. These conditions characterize the arithmetic mean divisor method and therefore establish (4). For objective function $\sum (q_i - a_i)^2/a_i$, we have

$$f_i(a_i - 1) - f_i(a_i) = \frac{q_i^2}{a_i(a_i - 1)} - 1$$

and

$$f_j(a_j) - f_j(a_j + 1) = \frac{q_j^2}{a_j(a_j + 1)} - 1,$$

so we must have $q_i^2/(a_i(a_i - 1)) \geq q_j^2/\big(a_j(a_j + 1)\big)$, or $q_i/d(a_i - 1) \geq q_j/d(a_j)$ with $d(a) = \sqrt{a(a + 1)}$, whenever $a_i \geq 1$ and $i \neq j$. These conditions characterize the geometric mean divisor method and therefore establish (1). For objective function $\sum q_i(\ln q_i - \ln a_i)$, we have

$$f_i(a_i - 1) - f_i(a_i) = q_i\big(\ln a_i - \ln(a_i - 1)\big)$$

and

$$f_j(a_j) - f_j(a_j + 1) = q_j\left(\ln(a_j + 1) - \ln a_j\right),$$

so we must have

$$q_i\left(\ln a_i - \ln(a_i - 1)\right) \geq q_j\left(\ln(a_j + 1) - \ln a_j\right),$$

or

$$\frac{q_i}{d(a_i - 1)} \geq \frac{q_j}{d(a_j)}$$

with $d(a) = 1/\ln\left((a + 1)/a\right)$, whenever $a_i \geq 1$ and $i \neq j$. These conditions characterize the logarithmic mean divisor method and therefore establish (2). For objective function $\sum a_i(\ln a_i - \ln q_i)$, we have

$$f_i(a_i - 1) - f_i(a_i) = (a_i - 1)\ln(a_i - 1) - a_i \ln a_i + \ln q_i$$

and

$$f_j(a_j) - f_j(a_j + 1) = a_j \ln a_j - (a_j + 1)\ln(a_j + 1) + \ln q_j,$$

so we must have

$$(a_i - 1)\ln(a_i - 1) - a_i \ln a_i + \ln q_i \geq a_j \ln a_j - (a_j + 1)\ln(a_j + 1) + \ln q_j,$$

or

$$\frac{q_i}{d(a_i - 1)} \geq \frac{q_j}{d(a_j)}$$

with $d(a) = (a+1)^{a+1}/(ea^a)$, whenever $a_i \geq 1$ and $i \neq j$. These conditions characterize the identric mean divisor method and therefore establish (3). ■

Note that objective functions (1) and (4) are mirror images in terms of the roles played by a_i and q_i. Likewise, objective functions (2) and (3) are mirror images. (1) and (4) represent weighted least-squares distances. (2) and (3) represent relative entropy distances (discussed shortly). We now explore the relationships between all four of these objective functions in the context of information theory and statistical hypothesis testing.

4. CONNECTIONS TO INFORMATION THEORY AND STATISTICS. In [**3**, p. 13], the entropy associated with a random variable with known finite probability distribution $\{u_1, \dots, u_n\}$ is defined as $H = -\sum u_i \log u_i$. Normally, the logarithm is base 2 and H is expressed in bits. An intuitive way to think about entropy is in the context of data compression [**3**, Chapter 5]. If you have a file consisting of a sequence of characters independently selected from a specified distribution, then entropy represents a lower bound on the average number of bits per character in the compressed file, or, what is the same, a lower bound on the expected binary codeword length. The entropy lower bound cannot be attained in all cases, but practical coding schemes approach it.

The Kullback-Leibler relative entropy distance measure [**3**, p. 18] between two finite probability distributions $\{u_1, \dots, u_n\}$ and $\{v_1, \dots, v_n\}$ is defined as

$$D = \sum u_i \log\left(\frac{u_i}{v_i}\right) = \sum u_i(\log u_i - \log v_i).$$

In the context of data compression, relative entropy measures the inefficiency associated with assuming the wrong character frequency distribution. If you use a file compression algorithm assuming distribution $\{v_1, \dots, v_n\}$ and apply it to data with actual distribution $\{u_1, \dots, u_n\}$, then $H + D$ bits represents a lower bound on the expected codeword length in the compressed file. Hence, the compressed file is larger than it would have been if you had used an optimal compression algorithm for your data.

Objective functions (2) and (3) are of the relative entropy form, except for a scale factor, considering distributions $\{a_i/h\}$ and $\{q_i/h\}$ and employing natural logarithms. Moreover, as indicated in [**3**, p. 333], (1) and (4) can be considered as Taylor approximations to (2) and (3), respectively. Consider the function $f_i(a_i) = a_i(\ln a_i - \ln q_i)$ and expand f_i about q_i so that

$$\begin{aligned} f_i(a_i) &= f_i(q_i) + f_i'(q_i)(a_i - q_i) + \frac{1}{2} f_i''(q_i)(a_i - q_i)^2 + \cdots \\ &= a_i - q_i + \frac{1}{2}(a_i - q_i)^2/q_i + \cdots . \end{aligned}$$

It follows that $\sum f_i(a_i) = \frac{1}{2}\sum (a_i - q_i)^2/q_i + \cdots$ or that (4) is a Taylor approximation to two times (3). Similarly, (1) is a Taylor approximation to two times (2) with the roles of a_i and q_i reversed.

We can get somewhat more insight on these comparisons within the theory of statistical hypothesis testing for multinomial trials [**5**, pp. 549–553]. Imagine an experiment with h independent trials where at each trial a ball is deposited in urn i with probability p_i/p, $i = 1, \ldots, s$. Let a_i denote the accumulated number of balls in urn i with expectation $q_i = hp_i/p$. In statistical theory, the likelihood ratio test statistic for conformance of the a_i outcomes with the q_i expectations is simply $2\sum a_i(\ln a_i - \ln q_i)$, i.e., two times (3). An older alternative is the classical Pearson test statistic $\sum(a_i - q_i)^2/q_i$, i.e., (4). In real experimental situations, both of these test statistics are asymptotically chi-square distributed with $s - 1$ degrees of freedom when the hypothesized probabilities and expectations are true, whereas improbably large values lead to a rejection of that hypothesis. The bottom line is that small values of these statistics provide credibility to the hypothesized parameters.

Now, we don't really have a statistical experiment, and the a_i are allocations, not random outcomes. Nevertheless, these connections lend support to (3) and (4) as natural objective functions for the apportionment problem, and in particular to Balinski and Young's preference for Webster over Hill. It certainly seems less intuitive to reverse roles and consider the q_i as outcomes and the a_i as expectations.

The two test statistics, likelihood ratio and Pearson, are asymptotically equivalent, but the former is generalized from the optimal Neyman-Pearson lemma ([**3**, pp. 304–309] and [**5**, pp. 431–433]), is more exact, and is generally favored by statisticians. This would tend to favor objective function (3) over (4). While this doesn't provide a very strong endorsement of (3), it does support the notion that the identric mean method is a logical contender to Webster, and moreover that both are preferable to Hill.

5. HISTORICAL APPORTIONMENTS. Appendix B in [**1**] contains comparative allocations for all decennial apportionments from 1790–2000 under existing divisor methods. The following table adds the logarithmic and identric methods compared to the geometric (Hill) and arithmetic (Webster) methods for 1920–1990, the modern era with house size 435. Divisor calculations incorporating unit lower bounds were straightforwardly executed in Excel. The focus here is only on those states where differences arose. The years 1930 and 2000 are excluded because the four methods produced identical allocations. As an aside, the minimal divisors associated with the geometric, logarithmic, identric, and arithmetic mean methods in 1990 were 574847.5, 574630.0, 574412.5, and 574109.8 respectively.

In 1990, the arithmetic method allocated one additional seat to Massachusetts and one less to Oklahoma as compared to the other three methods. Likewise, in 1980 the arithmetic method allocated one additional seat to Indiana and one less to New Mexico. 1970 is more complicated in that Connecticut got an additional seat from arithmetic, Oregon one less seat from geometric, Montana one less seat from arithmetic, and South Dakota one more seat from geometric. 1960 is similar to 1980 and 1990, with the arithmetic method allocating one additional seat to Massachusetts and one less seat to New Hampshire. In 1950, we have a different line-up where the arithmetic and identric methods allocated an additional seat to California and one less to Kansas. Similarly, in 1940 the arithmetic and identric methods allocated an additional seat to Michigan and one less to Arkansas. 1920 is the most complicated case of all. The arithmetic and identric methods agreed on the larger states, New York and North Carolina, and on the smaller states, New Mexico and Vermont, but they differed on Virginia and Rhode Island. Overall, the identric and arithmetic methods produced identical allocations in some instances (1930, 1940, 1950, 2000) but not in others (1920, 1960, 1970, 1980,

Table 2. Congressional allocations for house size 435

Year	State	Quota	Geometric	Logarithmic	Identric	Arithmetic
1990	Massachusetts	10.532	10	10	10	11
	Oklahoma	5.516	6	6	6	5
1980	Indiana	10.574	10	10	10	11
	New Mexico	2.504	3	3	3	2
1970	Connecticut	6.503	6	6	6	7
	Oregon	4.500	4	5	5	5
	Montana	1.496	2	2	2	1
	South Dakota	1.435	2	1	1	1
1960	Massachusetts	12.543	12	12	12	13
	New Hampshire	1.479	2	2	2	1
1950	California	30.722	30	30	31	31
	Kansas	5.529	6	6	5	5
1940	Michigan	17.453	17	17	18	18
	Arkansas	6.473	7	7	6	6
1920	New York	42.919	42	42	43	43
	North Carolina	10.581	10	10	11	11
	Virginia	9.547	9	9	9	10
	Rhode Island	2.499	3	3	3	2
	New Mexico	1.461	2	2	1	1
	Vermont	1.457	2	2	1	1

1990). The geometric and logarithmic methods were closely aligned, differing only in 1970.

REFERENCES

1. M. L. Balinski and H. P. Young, *Fair Representation: Meeting the Ideal of One Man, One Vote*, 2nd ed., Brookings Institution Press, Washington, DC, 2001.
2. M. Balinski, What is just?, this MONTHLY **112** (2005) 502–511.
3. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Interscience, New York, 1991.
4. G. Grimmett, Stochastic apportionment, this MONTHLY **111** (2004) 299–307.
5. P. W. Zehna, *Probability Distributions and Statistics*, Allyn and Bacon, Boston, 1970.

BOB AGNEW received his Ph.D. in applied mathematics from Northwestern in 1968 under Erhan Cinlar. He is currently a marketing analytics director for Discover Financial Services.
Discover Financial Services, Riverwoods, IL 60015-3851
raagnew@aol.com or robertagnew@discover.com