Proc. Natl. Acad. Sci. USA
Vol. 75, No. 8, pp. 3537–3539, August 1978
Applied Mathematical Sciences

Quota methods for congressional apportionment are still non-unique

(representation/fair division)

JOHN P. MAYBERRY

Department of Mathematics, Brock University, St. Catharines, Ontario, Canada L2S 3A1; and Department of Statistics, Princeton University, Princeton, New Jersey 08540

Communicated by John W. Tukey, April 26, 1978

ABSTRACT Balinski and Young described a "quota method" for congressional apportionment and recommended it as "the only method satisfying three essential axioms" [Balinski, M. L. & Young, H. P. (1974) *Proc. Natl. Acad. Sci. USA* 71, 4602–4606]. This paper points out and repairs a slight defect in one of those axioms, producing a quota method slightly different from that described previously. It also presents an alternative to the "consistency" axiom of the paper and describes the "dual quota" method, uniquely satisfying the alternative axioms (which have exactly as much justification as the originals).

1. The apportionment problem

The apportionment problem is to allocate the h seats of a legislature (or "house") among s states in proportion to their respective populations, $p_1, \ldots, p_s$, subject to overriding minima, $r_1, \ldots, r_s$, and maxima, $b_1, \ldots, b_s$. If each state could receive a nonintegral number of seats, an easy calculation (formalized in section 3 below) would specify the correct apportionment; the resulting numbers are called the *exact quotas* of the respective states. But each state must receive an integral number of seats, and therefore some suitable integers must be used to approximate the exact quotas. We may think of an apportionment method as an effective intepretation of the words "suitable" and "to approximate." Thanks to the clear distinction made by Balinski and Young (1, 2) between "solutions" and "methods," the present paper can largely disregard the rare but vexatious "ties," such as occur when two states have exactly equal populations.

2. The work of Balinski and Young

Balinski and Young (1) described an apportionment method, the "quota method," that satisfied three axioms intended to summarize the essential desiderata. The first axiom, "house-monotonicity," excludes the Alabama paradox (see section 5 below). The second axiom, "the quota condition," limits the discrepancy between the exact quota and any acceptable apportionment. The third axiom, "consistency," excludes capricious or discriminatory methods. They (1, 2) presented a precise and a very general definition of a "consistent" method as well as general background and further references. In fact, they stated (1) and proved (2) that the quota method is the unique method to satisfy all three axioms, even if minima (but not maxima) are imposed on the portions of the various states. Although the maxima specified in the Constitution will not influence the results of any known apportionment method for the present 435-seat Congress, they influenced Washington's decision to veto the first apportionment bill. I allow for both maxima and minima, to reveal the full duality described in section 4 below.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Note that this paper uses the word "consistent" as defined in refs. 1 and 2, and not as in refs. 3 and 4. The difference between those definitions explains an apparent inconsistency among the published theorems: "There exists no Huntington method [5] satisfying quota . . . ," and "There exists a unique consistent, house-monotone method [the quota method] satisfying quota," both on p. 4604 of ref. 1 and "An apportionment method M is [house-]monotone and consistent if and only if it is a Huntington method," p. 612 of ref. 4.

3. Notation and definitions

Bold-face letters denote s-tuples of real numbers indexed by i, where i is restricted to be one of $1, 2, \ldots, s$, and all summations are over all i. An apportionment *problem* is a set $(\mathbf{p},\mathbf{r},\mathbf{b},h)$ as above, with p_i, r_i, b_i, and h integral, $p_i > 0$, $0 \le r_i \le b_i$, and $\Sigma r_i = h_* \le h \le h^* = \Sigma b_i$. An *apportionment* for the problem $(\mathbf{p},\mathbf{r},\mathbf{b},h)$ is an s-tuple $\mathbf{a} = (a_1, \ldots, a_s)$ of integers called *portions*, with $r_i \le a_i \le b_i$ for each i, and $\Sigma a_i = h$.

An apportionment *solution* is a function $\mathbf{f}$, which to any such problem assigns an apportionment $\mathbf{a} = \mathbf{f}(\mathbf{p},\mathbf{r},\mathbf{b},h)$. [Note that the "pure" problem, without minima or maxima, is the special case $\mathbf{r} = (0, \ldots, 0)$ and $\mathbf{b} = (h, h, \ldots, h)$.] An apportionment *method* is a nonempty set of solutions. A method $\mathbf{M}$ is called *house-monotone* if, for every problem and any $\mathbf{f}$ in $\mathbf{M}$, $\mathbf{f}(\mathbf{p},\mathbf{r},\mathbf{b},h) \le \mathbf{f}(\mathbf{p},\mathbf{r},\mathbf{b},h + 1)$ unless $h \ge h^*$ so that the right side is undefined.

We notice that the exact quotas q_i satisfy $q_i = \max[r_i, \min(b_i, \lambda \cdot p_i)]$, where $\lambda = \lambda(h)$ is chosen so that $\Sigma q_i = h$. Since $g(\lambda) = \Sigma \max[r_i, \min(b_i, \lambda \cdot p_i)]$ is a continuous nondecreasing function of λ, with $g(0) = h_*$ and $g(\lambda) = h^*$ for large enough λ, the exact quotas are unique. *Lower quotas* and *upper quotas* are defined respectively by $\ell_i = \lfloor q_i \rfloor$ (the greatest integer not exceeding q_i) and $u_i = \lceil q_i \rceil$ (the least integer not less than q_i). An apportionment method $\mathbf{M}$ is said to *satisfy lower quota* if always $\mathbf{f}(\mathbf{p},\mathbf{r},\mathbf{b},h) \ge \boldsymbol{\ell}(\mathbf{p},\mathbf{r},\mathbf{b},h)$; to *satisfy upper quota* if always $\mathbf{f}(\mathbf{p},\mathbf{r},\mathbf{b},h) \le \mathbf{u}(\mathbf{p},\mathbf{r},\mathbf{b},h)$; and to *satisfy quota* if both conditions hold. Cogent arguments for requiring that a method satisfy quota, originally given by Daniel Webster, are quoted in ref. 2. Because the above definition of upper quota differs from that of ref. 1 in case $\mathbf{r} \ne 0$, a slightly modified "quota method" will result (see examples below).

4. Duality and the dual quota method

Once maxima as well as minima are considered in apportionment problems, a *duality* can be defined in which: maxima correspond with minima; Huntington's method (5) of smallest divisors corresponds with his method of greatest divisors; an upward induction (using house sizes increasing from h_* to h^*) corresponds with a downward induction; "greater than" corresponds with "less than"; lower quota corresponds with upper quota (as the latter is defined in this paper); and the method of major fractions corresponds with itself. Thus, for example, any proof that the method of smallest divisors satisfies upper quota

Table 1. Several apportionments for the "1984-B" population project of ref. 1

State	Population	Exact quotas*		Apportionments†					
		$q0$	$q1$	EP	BY	Q0	Q1	DQ0	DQ1
Alabama	3,608,877	7.099	7.088	7	7	7	7	7	7
Alaska	329,928	0.649	1.000	1	1	0	1	1	1
Arizona	1,885,014	3.708	3.702	4	4	4	4	4	4
Arkansas	1,948,052	3.832	3.826	4	4	4	4	4	4
California	21,944,556	43.167	43.102	41	44	44	44	43	43
Colorado	2,410,663	4.742	4.735	5	5	5	5	5	5
Connecticut	3,438,575	6.764	6.754	7	7	7	7	7	7
Delaware	802,199	1.578	1.576	2	1	1	1	2	2
Florida	7,671,182	15.090	15.067	15	15	15	15	15	15
Georgia	5,053,140	9.940	9.925	10	10	10	10	9	10
Hawaii	840,834	1.654	1.651	2	1	1	1	2	2
Idaho	804,232	1.582	1.580	2	1	1	1	2	2
Illinois	12,290,721	24.177	24.140	23	25	25	25	24	24
Indiana	5,570,655	10.958	10.941	11	11	11	11	10	10
Iowa	2,958,171	5.819	5.810	6	6	6	6	6	6
Kansas	2,456,416	4.832	4.825	5	5	5	5	5	5
Kentucky	3,465,010	6.816	6.806	7	7	7	7	7	7
Louisiana	3,968,799	7.807	7.795	8	8	8	8	7	7
Maine	1,298,870	2.555	2.551	3	2	2	2	3	3
Maryland	3,978,966	7.827	7.815	8	8	8	8	7	7
Massachusetts	6,086,136	11.972	11.954	12	12	12	12	11	11
Michigan	9,489,634	18.667	18.639	18	19	19	19	18	18
Minnesota	4,003,368	7.875	7.863	8	8	8	8	8	8
Mississippi	2,421,339	4.763	4.756	5	5	5	5	5	5
Missouri	5,108,552	10.049	10.034	10	10	10	10	10	10
Montana	773,730	1.522	1.520	2	1	1	1	2	2
Nebraska	1,834,178	3.608	3.603	4	3	4	3	4	4
Nevada	534,291	1.051	1.049	1	1	1	1	1	1
New Hampshire	842,868	1.658	1.655	2	1	1	1	2	2
New Jersey	7,676,299	15.100	15.077	15	15	15	15	15	15
New Mexico	1,313,105	2.583	2.579	3	2	2	2	3	3
New York	19,842,030	39.031	38.972	37	40	40	39	39	38
North Carolina	5,552,320	10.922	10.905	11	11	11	11	10	10
North Dakota	755,938	1.487	1.485	2	1	1	1	2	2
Ohio	11,735,587	23.085	23.050	22	23	23	23	23	23
Oklahoma	2,901,743	5.708	5.699	6	6	6	6	6	6
Oregon	2,385,753	4.693	4.686	5	5	5	5	5	5
Pennsylvania	12,779,259	25.138	25.100	24	26	26	26	25	25
Rhode Island	1,316,663	2.590	2.586	3	2	2	2	3	3
South Carolina	2,905,301	5.715	5.706	6	6	6	6	6	6
South Dakota	752,887	1.481	1.479	2	1	1	1	2	2
Tennessee	4,031,836	7.931	7.919	8	8	8	8	8	8
Texas	12,228,700	24.055	24.019	23	24	24	24	24	24
Utah	1,360,383	2.676	2.672	3	2	2	2	3	3
Vermont	485,488	0.955	1.000	1	1	1	1	1	1
Virginia	5,026,197	9.887	9.872	10	10	10	10	9	9
Washington	3,519,403	6.923	6.913	7	7	7	7	7	7
West Virginia	1,854,004	3.647	3.641	4	3	4	4	4	4
Wisconsin	4,519,866	8.891	8.878	9	9	9	9	8	8
Wyoming	376,698	0.741	1.000	1	1	0	1	1	1

* $q0$, exact quotas with minima $r_i = 0$; $q1$, exact quotas with minima $r_i = 1$.

† EP, equal proportions; BY, the Balinski–Young "quota" method with $r_i = 1$ (both from ref. 1); Q0, the quota method of this paper, $r_i = 0$; Q1, same, $r_i = 1$; DQ0, the dual quota method, $r_i = 0$; DQ1, same, $r_i = 1$.

can be translated (by a purely mechanical process) into a proof that the method of greatest divisors satisfies lower quota.

Under this duality, the quota method of refs. 1 and 2—defining "upper quota" as above—will correspond to another algorithm, to be called the dual quota method, which will have exactly as much basis for acceptability as the quota method itself. Similarly, the proof that the quota method is the unique method satisfying the three axioms of Balinski and Young translates into a proof that the dual quota method is the unique method satisfying three equally reasonable axioms.

Table 1 shows that the quota and dual quota solutions may differ—in as many as 24 of the 50 states in the 1984B example from ref. 1. This situation arises because the "consistency" condition of ref. 1 is not self-dual—the dual quota method satisfies a consistency condition dual to that satisfied by the quota method. Other monotone methods have been found that satisfy quota; it happens that quota and dual quota are the easiest to define and to compute (see section 8 below).

5. The Alabama paradox

The importance of house-monotonicity can be seen by considering the "Vinton" or "Hamilton" method—viz: "Give each state its lower quota, and one more seat to each of the $h - \Sigma \ell_i$ states with greatest remainders $q_i - \ell_i$." The pure problem with $\mathbf{p} = (1,3,3)$ then gives the apportionments (1,1,1) with house size 3, and (0,2,2) with house size 4; the first state *loses* a seat when the house *increases* from 3 to 4. Much colorful discussion has made clear that the Congress finds this "Alabama paradox" (named in honor of its first victim) unacceptable.

6. Consistency

We express in our notation some salient portions of the definition of *consistency,* from p. 4604 of ref. 1.

Any house-monotone solution f may be characterized by identifying, for any **p**, **r**, and **b**, the sequence of states that gain seats as the house increases from h_* to $h_* + 1, \ldots, h^*$. The *eligible set at* h is the set $E(h) = \{i \mid f_i(h-1) < q_i(h)\}$ of states that could receive the h'th seat without violating upper quota; $f_i(h-1)$ is the *previous portion* of state i.

A house-monotone method **M** is called *consistent* if the choice of state to receive the h'th seat is governed entirely by priority among the eligible states, where relative priority between any two states is determined only by their populations and (immediately) previous portions. Without completing the details, it is clear that this consistency condition explicitly protects the solution from violating upper quota but not from violating lower quota; it is thus rather natural that the only consistent method that satisfies quota is an analog of greatest divisors—which intrinsically satisfies lower quota. [Although "natural," that uniqueness theorem—proven in ref. 2—is far from trivial; it is relatively straightforward to modify that proof to allow maxima as well as minima, and to incorporate the definition of upper quota given above.]

7. The dual quota method

The concepts of dual-eligibility and dual-consistent can be derived from the above definitions; if we consider the sequence of states that *lose* seats as the house *decreases* from h^* to $h^* - 1, \ldots, h_*$, we define the *dual-eligible set at* h as the set $E'(h) = \{i \mid f_i(h+1) > q_i(h)\}$ of states that could lose the $(h+1)$th seat without violating lower quota. We then define a solution f as *dual-consistent* if the choice of the losing state is governed by priority within the dual-eligible set, where relative priority of two states is determined by their populations and previous (i.e., at the next-higher house-size) portions. It is natural that in this case an analog of the method of smallest divisors (which intrinsically satisfies upper quota) has the desired properties. The modified proof mentioned above may be translated mechanically into a proof of the following theorem.

THEOREM. *There exists a unique dual-consistent house-monotone method satisfying quota*. That method, called the dual-quota method, is the set of all solutions ψ obtained recursively as follows: *(i)* $\psi_i(\mathbf{p},\mathbf{r},\mathbf{b},h^*) = b_i$ for all i; *(ii)* Given $a_i = \psi_i(\mathbf{p},\mathbf{r},\mathbf{b},h)$ and $E'(h-1) = \{i \mid a_i > q_i(h-1)\}$, let k be a state in $E'(h-1)$ such that $p_k/(a_k-1) \leq p_i/(a_i-1)$ for all i in $E'(h-1)$. Then $\psi_k(h-1) = a_k - 1$, $\psi_i(h-1) = a_i$ for $i \neq k$.

8. Other house-monotone methods satisfying quota

Instead of replacing the "consistency" condition of ref. 1 by its dual, we might modify the definition of "eligible set" (and thus of "consistency") by defining a state to be eligible at h if that state could receive the next (h'th) seat without causing the resulting apportionment to violate *either upper or lower* quota. Clearly, the quota method is still consistent under this definition; examples have shown that the dual quota method is not, although it would satisfy the *dual* of the new condition. It is uncertain whether the quota method would remain the unique method under this new definition; at any rate, the uniqueness-proof given in ref. 2 certainly is invalidated.

As mentioned in section 4, other methods have been discovered that are monotone and satisfy quota; many of those are analogs of previously known methods (equal proportions, major fractions, or Hamilton/Vinton), just as quota and dual quota are analogs of greatest divisors and smallest divisors, respectively. (Two papers discussing those methods will be published elsewhere by Jonathan W. Still and by J.P.M.)

9. Selection of the method for apportionment of Congress

Many of the desiderata for an apportionment method seem so natural that one would not think to impose them as requirements—until examples are found that violate them. (That occurred with the quota requirement in 1832 and with monotonicity in 1881.) Other formal conditions may be found, as well as the informal conditions of "simplicity," "computational ease," and "impartiality between large and small states." It seems appropriate to define exact quotas for the constrained case and then to impose the quota condition; it also seems desirable that the number of persons represented by a congressman be nearly constant from one state to another. Among monotone methods, the first condition would militate in favor of one of the quota methods, whereas the second would favor the method of equal proportions (5). The Congress may wish to focus attention on the question, Should the relative deviations from exact quota, which will necessarily be large for the smaller states, justify similar relative deviations from exact quota for the larger states, where they will generally not be necessary? In the 1984B hypothetical census figures of ref. 1, fractional parts of exact quotas for most of the small states exceed 0.5; thus, in the equal proportions solution, which emphasizes *relative* discrepancies, many small states gain a fraction—at the expense of the largest states, six of which obtain less than their lower quotas. (The 1984A example from ref. 1 shows the opposite phenomenon—many small states lose a fraction and some large states exceed their upper quotas.) Huntington's analysis (and the Congress's acceptance of that analysis) constitutes a strong case for the equal proportions solution, but Huntington might well have given more weight to satisfying quota if he had known that it was compatible with the Congressional insistence on house-monotonicity. Webster's arguments (quoted in ref. 2) imply that a nonquota apportionment is unconstitutional; that viewpoint was apparently not explored mathematically until refs. 1 and 2 were published.

If a method satisfying quota is to be selected, should it be the quota method or the dual quota method, or some other (see section 8)? Quota inherits from greatest divisors the tendency to favor large states, whereas dual quota inherits from smallest divisors the tendency to favor small states; all the other quota methods are in that respect "between" these two. This author's preference would be for the quota analog of major fractions, which happens to be self-dual; good cases can also be made for the quota analog of the Hamilton–Vinton method and for the quota analog of equal proportions. Any one of these three would be an attractive compromise between quota and dual quota, much as equal proportions was an attractive compromise between greatest divisors and smallest divisors.

1. Balinski, M. L. & Young, H. P. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 4602–4606.
2. Balinski, M. L. & Young, H. P. (1975) *Am. Math. Mon.* **82,** 701–730.
3. Balinski, M. L. & Young, H. P. (1977) *Am. Math. Mon.* **84,** pp. 450–455.
4. Balinski, M. L. & Young, H. P. (1977) *SIAM J. Appl. Math.* **33,** 607–618.
5. Huntington, E. V. (1928) *Am. Math. Soc. Trans.* **30,** 85–110.