The proportional representation problem in the Second Chamber: an approach via minimal distances

by H. J. J. te Riele*

Abstract The purpose of this paper is to discuss the proportional representation problem in the Second Chamber of Dutch Parliament. Firstly, the present procedure for solving this problem is described, together with a possible alternative procedure, recently proposed by F. J. Lisman. Next, the problem is formulated as that of minimizing some distance coefficient between the distribution of the votes of the electorate and that of the seats in the Second Chamber. For distances which satisfy a specific convexity condition a simple and straight forward algorithm is given for computing a distance minimizing seat distribution. The procedure of Lisman is shown to have three attractive properties by which it is distinguished from the other usual procedures for solving the proportional representation problem. Finally, the principle of the weighted vote is introduced as a means of breaking the deadlock which always comes into being, because of the inevitable discrepancy between the vote distribution and the actual seat distribution.

1 Introduction

Each four years new representatives in the Second Chamber of Dutch Parliament have to be elected. Political parties then nominate their candidates and the numbers of votes for these parties have to be translated into the numbers of corresponding representatives (or seats), which sum up to 150. According to the Constitution of the Netherlands the numbers of seats shall be in the same proportion as the numbers of votes. However, in general this proportionality cannot be realized exactly by whole numbers of seats, so that a seat distribution should be found which fits the vote distribution "as closely as possible", in some sense. This problem is known as the Proportional Representation problem (abbreviated: the PR-problem).

In order to be able to quantify the vague concept of closeness, we shall have to define the concept of a "distance" between seat distributions; a "solution" of the PR-problem will then be a seat distribution which has minimal distance to the so-called theoretical seat distribution (to be defined later). The problem of computing this solution can be formulated as a nonlinear integer programming problem. Algorithms for its solution, together with computer implementations, are available. However, these algorithms are designed for very general programming problems, and we recommend another algorithm, which exploits a specific convexity property of the distances in question.

Now, of course, different distances will generally yield different solutions of the PR-problem, and the question arises whether a distance exists which leads to a "fairest" or a "best" solution. Obviously, the answer is negative because the concept of fairness will depend on the way we look at the PR-problem: e.g., should the parties be treated as fairly as possible, or the individual voters? Different starting points will generally lead to different distances, so no compelling distance concept exists.

* Mathematical Centre, Amsterdam.

The scientist studying politics, assisted by the mathematician, should try to formulate the kind of fairness to be aimed at. Next, this should be translated into a distance concept. This interplay between political sciences and mathematics might be called "politicometrics", analogous to disciplines like econometrics, sociometrics, psychometrics and jurimetrics. Once the distance concept is decided upon, it is the task and the responsability of the mathematician, to be assisted by the scientist studying politics, to analyze the mathematical properties of the procedure for solving the PR-problem, based on minimizing this given distance function.

In Section 2 we describe the present procedure for the apportionment of representatives in the Second Chamber of Dutch Parliament, known as the procedure of HAGENBACH-BISCHOFF, abbreviated HB. We quantify the frequently raised objection to this procedure, viz., that it favours the greater parties. Section 3 refers to a possible alternative procedure, proposed in 1973 by F. J. LISMAN [3, 4], who called it the procedure RE, which stands for "Rounded off Exactly". In Section 4 we define seven different distances. A very simple and efficient algorithm is given for computing the distance minimizing seat distribution for each of these seven distances. In Section 5 it is shown that HB corresponds to the distance no. 3 and that RE corresponds to no. 5. Moreover, three special properties of RE are described and compared with properties of some other common procedures. In Section 6, finally, we pay attention to the possibility of assigning so-called weighted votes. This has the effect that the proportions of the votes of the electorate are perfectly preserved in the (weighted) votes of the representatives in Parliament, no matter the procedure used for the determination of the actual seat distribution in Parliament.

Politicians may find a non-mathematical survey with respect to the PR-problem in LISMAN, LISMAN and TE RIELE [5].

We adopt the following notations. There are m parties, numbered $1, 2, \ldots, m$, which have received $v_1, v_2, \ldots, v_m$ valid votes, respectively, with $v_1 \geqq v_2 \geqq \ldots \geqq v_m > 0$ and $V_i := v_1 + v_2 + \ldots + v_i$, $1 \leqq i \leqq m$. The total number of seats to be distributed is S. Let $Q_m := V_m/S$, and let α be some fixed positive real number. In many allocation procedures those parties are immediately excluded, for which $v_i < \alpha Q_m$. Examples are $\alpha = 1$ for the Dutch Second Chamber and $\alpha = 0.75$ for the Dutch Municipal Councils. We assume that n parties, $n \leqq m$, have passed this "voting-threshold" sieve. Now we formulate the PR-problem as follows: For any given vote distribution $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ a seat distribution $\boldsymbol{s} = (s_1, s_2, \ldots, s_n)$ has to be determined such that $s_1 + s_2 + \ldots + s_n = S$, $s_i \in \mathbb{N}_0 (= \mathbb{N} \cup \{0\})$, and such that it fits a given theoretical seat distribution $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)$ "as closely as possible". Now an obvious way to define the theoretical seat distribution $\boldsymbol{r}$ is by choosing $r_i := v_i S/V_n$, so that $\sum_{i=1}^{n} r_i = S$. However, sometimes we shall also have to work with $r_i := v_i S/V_m$, so that $\sum_{i=1}^{n} r_i \leqq S$, with equality if and only if $n = m$.

2 The present procedure in the Second Chamber of Dutch Parliament

The present procedure for the allocation of seats in the Second Chamber of Dutch

Parliament is known as the procedure of HAGENBACH-BISCHOFF and it is in use since 1933. It is essentially equivalent with the method of D'HONDT [3], and known in the American literature as the method of the Greatest Divisors [1, 2]. It can be described as follows.

The total number of representatives S amounts to 150. The total number of valid votes divided by 150 is called the quota (Q_m). At present there is a so-called voting-threshold, which means that a party should have scored at least Q_m votes in order to receive a seat. Now the total number of votes of each admitted party is divided by Q_m (yielding, in our notation, the theoretical seat distribution $\boldsymbol{r}$ with $r_i = v_i S/V_m$), and the integer parts of the resulting numbers show how many seats are allocated initially to each party. Obviously, a small number of remaining seats is then left to be allocated. This is done as follows (we cite from the Electoral Law): "Successively every time one of the remaining seats is allocated to the party which, after the allocation, shows the greatest number of votes per seat".

A simple example with five parties and $S = 20$ may enlighten this procedure (see table 1). The second column of table 1 gives the numbers of valid votes v_i. The total number of valid votes is 1000, so $Q_5 = 1000/20 = 50$. Columns 3–8 give the relevant data leading to the final seat distribution $\boldsymbol{s} = (11, 4, 4, 1, 0)$. Intermediate seat distributions are denoted by s_i' and s_i''.

Table 1. The procedure HB applied to 5 parties with $S = 20$

party i	v_i	$v_i/50$	$s_i' = [v_i/50]$	$v_i/(s_i'+1)$	s_i''	$v_i/(s_i''+1)$	s_i
1	528	10.56	10	48	11	44	11
2	205	4.10	4	41	4	41	4
3	180	3.60	3	45	3	45	4
4	84	1.68	1	42	1	42	1
5	$3(< Q_5)$						
	$1000 = V_5$		18		19		20

The most important and fundamental objection to this procedure is that it favours the greater parties. This can be seen from table 1, but another example makes things still more clear. For $\boldsymbol{v} = (900, 94, 6)$ and $S = 20$, the procedure HB yields $\boldsymbol{s} = (19, 1, 0)$. Here the proportion of the votes v_1/v_2 is $900/94 = 9.4$, whereas the proportion of the seats s_1/s_2 is $19/1 = 19.0$. The seat distribution $\boldsymbol{s} = (18, 2, 0)$ looks much more "fair".

An impression of the extent to which greater parties are favoured by HB is given by the following simple observations. Let v_i be the number of votes of the i-th admitted party. Initially, this party gets $s_i' = [v_i/Q_m]$ seats (by $[x]$ we mean the greatest integer not exceeding x). We write

$$v_i = s_i' Q_m + e_i, \qquad 0 \leqslant e_i < Q_m. \tag{2.1}$$

The first remaining seat will be allocated to the party with greatest value of $v_i/(s_i'+1)$.

Now we compare two parties i and j. If $v_i/(s_i'+1) > v_j/(s_j'+1)$ then party j certainly does not get the first remaining seat. Eliminating s_i' and s_j' by use of (2.1) yields

$$e_j < Q_m(1-v_j/v_i)+e_iv_j/v_i. \tag{2.2}$$

As an example, put $v_j/v_i = 0.1$. Then (2.2) means that the smaller party j certainly does not get the first remaining seat, if its number of remaining votes e_j is smaller than 90% of Q_m, irrespective of the number of remaining votes e_i of the greater party i. The (19, 1, 0)-example above was constructed by using these observations. More "realistic" examples could also be easily constructed. So the procedure HB shows a clear bias in favour of the greater parties.

For the sake of completeness it should be added that recently HB was extended with the possibility for parties to make combinations. This means that initially parties belonging to the same combination are considered as one party. After the allocation of seats according to the procedure described above, the seats of a combination party are distributed among the different parties of the combination according to the so-called method of the Greatest Remainders (see Section 5). In certain cases a smaller party can get one extra seat by forming a combination with a greater (congenial) party, but, nevertheless, this combination possibility is not a completely satisfactory remedy for the bias of HB. Moreover, it is of an ad-hoc character, and it unnecessarily complicates the procedure as a whole.

Remark

There is another method, known in the American literature as the method of the Smallest Divisors [1, 2], abbreviated SD, which is the counterpart of HB in that it systematically favours the smaller parties. It can be described as follows. After the exclusion of the parties whose numbers of votes are less than the voting-threshold, the numbers $r_i = v_iS/V_m$ are rounded upwards. The resulting numbers show how many seats are allocated initially to each party. If their total exceeds S, then successively each party gives in one seat, which shows, after the surrender of the seat, the smallest number of votes per seat. If their total is less than S, then successively each party receives one seat, which shows, before the allocation, the greatest number of votes per seat. The systematic bias of SD in favour of the smaller parties can be shown in essentially the same way, as the bias of HB in favour of the greater parties was shown.

3 The procedure RE

Another procedure, called the procedure RE (which stands for "Rounded off Exactly"), was introduced in the Netherlands in 1973 by F. J. LISMAN. In a different form this procedure was proposed already in 1916 by WILLCOX [9] in the United States, where it is also known as the method of the Major Fractions [1, 2]. The so-called modified LAGUE-method (cf. [6]), used in Sweden, is essentially equivalent with RE. LISMAN describes his procedure in four points.

i. First of all those parties are left out, whose numbers of votes are less than the quota Q_m.
ii. Next, the number of votes of each admitted party is divided by the total number of votes of the admitted parties (thus totally excluding those parties whose numbers of votes are less than Q_m) and multiplied by S. The resulting numbers, generally no integers, show a seat distribution which corresponds exactly to the distribution of the votes. LISMAN calls it the "exact" distribution (in our notation, the theoretical seat distribution $\boldsymbol{r}$ with $r_i = v_i S/V_n$).
iii. Now the figures of the exact distribution are rounded off to integers, the fractional parts $\geqslant 0.5$ upwards, and < 0.5 downwards. If the rounded numbers sum up to S, then this is the desired distribution of the seats.
iv. It may happen that the sum of the rounded numbers (seats) is slightly larger or smaller than S. In such a case, the denominator in step ii. has to be increased or decreased a little bit, respectively, such that, after rounding off, the numbers of seats sum up to S. Except in the case of ties, this denominator can always be found. The probability of the occurrence of ties is extremely small in realistic cases, therefore we leave this complication out of discussion.

LISMAN points out the facts that during the process the mutual proportions of the figures remain the same and that the deviation from S, for which step iv. corrects, is smaller than the number of remaining seats, to be distributed in the procedure HB (but cf. our remark in Section 5 after lemma 5.1).

The procedure RE may be demonstrated by the following example (see table 2) with five parties and $S = 20$. We start with the same data as table 1. The parties 1–4 are admitted. The third column shows the "exact" seat distribution. Rounding off yields a sum of 21 seats, and replacing 997 by 1010 in step ii. yields, after rounding off, 20 seats. The second example of section 2 yields, with RE, the seat distribution (18, 2, 0).

Table 2. The procedure RE applied to 5 parties with $S = 20$

party i	v_i	$v_i \cdot \frac{20}{997}$	rounded off	$v_i \cdot \frac{20}{1010}$	rounded off s_i
1	528	10.59	11	10.46	10
2	205	4.11	4	4.06	4
3	180	3.61	4	3.56	4
4	84	1.69	2	1.66	2
	$997 = V_4$		21		20
5	$3 (< Q_5)$				

Comparing the results of the procedures HB and RE it appears that for our two examples RE produces results which fit the theoretical seat distribution better than HB. We return to this in the next sections.

As a final remark it may be stated that in the procedure HB votes for parties which are not admitted to Parliament flow over to a certain extent to the greater parties. In RE they have no function at all. In order to create a destination for these lost votes LISMAN considers it appropriate to offer the electors the opportunity to present a second vote, referring to the second party in case the first party is not admitted.

As a final (realistic) example table 3 gives the results of the elections in May 1977, both for HB and RE. The shift to the smaller parties is clear (parties 1 and 3 lose a seat to parties 5 and 6).

Table 3. HB and RE applied to the results of the elections of May 1977

			procedure HB			procedure RE		
i (party)	v_i		r_i	s_i		r_i	s_i	
1*	3,032,675=		54.67	57=		55.73	56=	
(PvdA+		2,813,793+			53+			52+
PPR+		140,910+			3+			3+
PSP)		77,972			1			1
2 (CDA)	2,655,391		47.87	49		48.80	49	
3 (VVD)	1,492,689		26.91	28		27.43	27	
4 (D'66)	452,423		8.16	8		8.31	8	
5*	256,431=		4.62	4=		4.71	5=	
(SGP+		177,010			3+			3+
GPV)		79,421			1			2
6 (CPN)	143,481		2.59	2		2.64	3	
7 (BP)	69,914		1.26	1		1.28	1	
8 (DS'70)	59,487		1.07	1		1.09	1	
	8,162,491		147.15	150		149.99	150	
Others	158,234							
	8,320,725							

* Parties 1 and 5 are combination parties; within a combination party the seats are distributed in HB according to the method of the greatest remainders, in RE according to RE.

4 Distance minimizing procedures

In the previous two sections we have described two procedures for solving the PR-problem in the form of a rule for calculating a seat distribution $\boldsymbol{s}=(s_1, s_2, \ldots, s_n)$, given a vote distribution $\boldsymbol{v}=(v_1, v_2, \ldots, v_n)$, subject to the restrictions $\Sigma s_i = S$ and $s_i \in \mathbb{N}_0$.

In this and following sections we adopt a different point of view. We define a distance function $d(.,.)$ of two parameters $\boldsymbol{r}$ and $\boldsymbol{s}$ (in this order), where $\boldsymbol{r}$ will always stand for the theoretical seat distribution; a solution of the PR-problem is now defined to be a seat distribution $\boldsymbol{s}$ which has *minimal distance* to the given theoretical seat distribution $\boldsymbol{r}$. In general, this solution exists and is unique, except in the case of ties, which occurs very rarely in practice (e.g., when two parties with the same number of votes have to share an odd number of seats).

In table 4 a collection of seven different distances is given, so we have seven dif-

Table 4. Seven distances $d(\boldsymbol{r}, \boldsymbol{s})$

No.	distance function $d(\boldsymbol{r}, \boldsymbol{s}) = \sum_{i=1}^{n} f_i(s_i)$	$f_i(x)$	value of x for which $f_i(x)$ is minimal
1	$\sum_{i=1}^{n} \lvert s_i - r_i\rvert^a \quad (a \geqslant 1)$	$\lvert x - r_i\rvert^a$	r_i
2	$\sum_{i=1}^{n} \lvert s_i/r_i - 1\rvert^a \quad (a \geqslant 1)$	$\lvert x/r_i - 1\rvert^a$	r
3	$\sum_{i=1}^{n} (s_i - r_i + \frac{1}{2})^2/r_i$	$(x - r_i + \frac{1}{2})^2/r_i$	$r_i - \frac{1}{2}$
4	$\sum_{i=1}^{n} (s_i - r_i - \frac{1}{2})^2/r_i$	$(x - r_i - \frac{1}{2})^2/r_i$	$r_i + \frac{1}{2}$
5	$\sum_{i=1}^{n} (s_i - r_i)^2/r_i$	$(x - r_i)^2/r_i$	r
6	$\sum_{i=1}^{n} (s_i - r_i)^2/s_i$	$(x - r_i)^2/x$	r_i
7	$\sum_{i=1}^{n} s_i \log(s_i/r_i)$	$x \cdot \log(x/r_i)$	r_i/e

ferent distance minimizing procedures for the solution of the PR-problem. The third column of table 4 lists the functions $f_i(\cdot)$ such that

$$d(\boldsymbol{r}, \boldsymbol{s}) = \sum_{i=1}^{n} f_i(s_i). \tag{4.1}$$

The last column gives the value of x for which $f_i(x)$ assumes its minimum.

It should be noted that only the distances nos. 1 and 2 are real distances in the sense that the triangle inequality holds. For $a = 2$, no. 1 is the square of the Euclidean distance between the vectors $\boldsymbol{r}$ and $\boldsymbol{s}$, whereas no. 2 is the square of the Euclidean distance between the vectors $(s_1/r_1, s_2/r_2, \ldots, s_n/r_n)$ and $(1, 1, \ldots, 1)$. The other five distances are neither symmetric in $\boldsymbol{r}$ and $\boldsymbol{s}$, nor do they satisfy the triangle inequality. Yet, for the uniformity and the ease of reasoning we have chosen to use the term "distance". We have chosen for these seven, and no other, distances because i. nos. 3–7 play a role in the rest of this paper, ii. nos. 1 and 2 naturally arise in certain reasonings for finding a "fair" solution of the PR-problem and iii. we wanted to avoid, if possible, distances which become infinite for $s_i = 0$, like $\Sigma|r_i/s_i - 1|^a$ and $\Sigma r_i \log(r_i/s_i)$.

When looking for a seat distribution $\boldsymbol{s}$ with minimal distance to a given theoretical seat distribution $\boldsymbol{r}$, one should realize that the distances nos. 3–7 share the following important property: if s has minimal distance to r, then it has also minimal distance to $cr=(cr_1, cr_2, \ldots, cr_n)$, for any positive real constant c (for distance no. 3, e.g., this follows from the observation that, if

$$\Sigma(s_i+\tfrac{1}{2}-r_i)^2/r_i = \Sigma(s_i+\tfrac{1}{2})^2/r_i - 2\Sigma(s_i+\tfrac{1}{2})+\Sigma r_i$$

is minimal, then also

$$c^{-1}\Sigma(s_i+\tfrac{1}{2})^2/r_i - 2\Sigma(s_i+\tfrac{1}{2})+c\Sigma r_i = \Sigma(s_i+\tfrac{1}{2}-cr_i)^2/(cr_i)$$

is minimal). Therefore, for distances nos. 3–7 we can always rescale r such that $\Sigma r_i = S$. On the other hand, for the distances nos. 1 and 2 we must carefully distinguish between the cases $r_i = v_iS/V_n$ and $r_i = v_iS/V_m$. This difference between nos. 1–2 and nos. 3–7 finds expression in the so-called "ALABAMA-PARADOX" which can occur for nos. 1–2, but not for nos. 3–7. We return to this point in Section 5.

A simple and efficient algorithm for computing the distance minimizing seat distribution is based on the following lemma. Let for given distance d and theoretical seat distribution r the symbol $\sigma_S(d, r)$ denote the set of all seat distributions s which have minimal distance to r, subject to $s_i \in \mathbb{N}_0$, $\Sigma s_i = S$, i.e.,

$$\sigma_S(d, r) := \{s=(s_1, s_2, \ldots, s_n) | s_i \in \mathbb{N}_0, \Sigma s_i = S, d(r, s) = \inf_{t_i \in N_0, \Sigma t_i = S} d(r, t)\}.$$

Lemma 4.1
Suppose that all functions $f_i(.)$ in (4.1), $i=1, 2, \ldots, n$, are convex. If $s \in \sigma_S(d, r)$, then there exists $s^* \in \sigma_{S+1}(d, r)$ and ${}^*s \in \sigma_{S-1}(d, r)$ with $s_i^* \geqq s_i$ $(i=1, 2, \ldots, n)$ and ${}^*s_i \leqq s_i$ $(i=1, 2, \ldots, n)$.

The proof of this lemma is given in Appendix A. One easily verifies that for all distances in table 4 all functions $f_i(.)$ are convex, so this lemma can be used, once we have found an element $s \in \sigma_S$, for some S. Therefore, lemma 4.1 suggests the following algorithm:

Algorithm for computing $s \in \sigma_S(d, r)$, for a given distance d from table 4, a theoretical seat distribution r and a total number of seats S.

STEP 1. Use column 4 in table 4 in order to find $s=(s_1, s_2, \ldots, s_n)$ such that every individual term $f_i(s_i)$ in (4.1) is as small as possible. If $\Sigma s_i = S$ we are ready. If $\Sigma s_i < S$, perform step 2, otherwise step 3.

STEP 2 ($\Sigma s_i < S$). Determine an index $i=i_0$ such that $f_i(s_i+1)-f_i(s_i)$ is as small as possible. Put $s_{i_0} := s_{i_0}+1$. Repeat step 2 until $\Sigma s_i = S$.

STEP 3 ($\Sigma s_i > S$). Determine an index $i=i_0$ such that $f_i(s_i-1)-f_i(s_i)$ is as small as possible. Put $s_{i_0} := s_{i_0}-1$. Repeat step 3 until $\Sigma s_i = S$.

Since for the distances nos. 1–5 the functions $f_i(.)$ are symmetric around their minima, it follows that step 1 will deliver for s_i: the nearest integer to the x for which $f_i(x)$ is minimal. So step 1 delivers for nos. 1, 2 and 5 that $s_i = \text{round}(r_i) = [x+0.5]$, for no. 3 that $s_i = [r_i]$ and for no. 4 that $s_i = [r_i]+1$. However, for nos. 6 and 7 the

Table 5a. Results of the algorithm for distances Nos. 1–5; $r = (9.061, 7.173, 5.265, 3.319, 1.182)$, $S = 26$

		s_i for distance No.				
i	r_i	1	2	3	4	5
1	9.061	9	9	10	9	9
2	7.173	7	8	7	7	7
3	5.265	5	5	5	5	6
4	3.319	4	3	3	3	3
5	1.182	1	1	1	2	1

functions $f_i(.)$ are not symmetric around their minima, so both integer values around the minima must be checked, in order to find the s_i in step 1 of the algorithm.

As an illustration we give in table 5a the results of applying the algorithm for distances nos. 1–5 to the example $r = (9.061, 7.173, 5.265, 3.319, 1.182)$ and $S = 26$. For nos. 1, 2, 3 and 5 step 1 delivers $s = (9, 7, 5, 3, 1)$ with $\Sigma s_i = 25$, hence only one "correction" with step 2 is needed. For no. 4 step 1 delivers $s = (10, 8, 6, 4, 2)$ with $\Sigma s_i = 30$, hence four corrections with step 3 are needed. Notice that r was chosen in such a way that the five distances yield five different seat distributions.

As a second illustration, table 5b lists the results of applying the algorithm for distances nos. 5–7 to the example $r = (5.496, 4.496, 3.710, 3.490, 2.808)$ and $S = 20$.

Table 5b. Results of the algorithm for distances Nos. 5–7; $r = (5.496, 4.496, 3.710, 3.490, 2.808)$, $S = 20$

		s_i for distance No.		
i	r_i	5	6	7
1	5.496	6	5	5
2	4.496	4	4	5
3	3.710	4	4	4
4	3.490	3	4	3
5	2.808	3	3	3

For no. 5 step 1 delivers $s = (5, 4, 4, 3, 3)$ with $\Sigma s_i = 19$, so step 2 is performed once. For no. 6 step 1 delivers $s = (6, 5, 4, 4, 3)$ with $\Sigma s_i = 22$, so step 3 is performed twice. For no. 7, finally, step 1 delivers $s = (2, 2, 1, 1, 1)$, so here step 2 must be performed 13 times. Notice that again r was chosen such that the three different distances have three different minimizing seat distributions. It was not easy to construct this example, since the distances nos. 5 and 6 are very good approximations of no. 7. Very often the seat distributions of nos. 5, 6 and 7 will coincide. We return to this in the next section.

5 A closer look at RE and some other procedures

In this section we shall describe three special properties of the procedure RE, and we shall also pay attention to properties of some other procedures. For shortness, by "the procedure d_i" we mean the procedure for solving the PR-problem by minimizing distance no. i of the actual seat distribution to the theoretical seat distribution ($1 \leqq i \leqq 7$). We shall need the following lemma.

Lemma 5.1

a. HB *minimizes the distance no.* 3,

b. SD *minimizes the distance no.* 4,

c. RE *minimizes the distance no.* 5.

Parts a. and c. of lemma 5.1 are proved in the Appendix B and the Appendix C, respectively. The proof of part b. can be safely left to the reader who has worked through the proof of part a.

Remark

As noticed already in the text before lemma 4.1, before we minimize one of the distances nos. 3–7 to $\boldsymbol{r}$, we can rescale $\boldsymbol{r}$ without affecting the resulting minimizing seat distribution $\boldsymbol{s}$. Therefore, lemma 5.1a immediately suggests an improvement in the present procedure HB: instead of working with $r_i = v_i/Q_m = v_i S/V_m$, one should use $r_i = v_i/Q_n = v_i S/V_n$. Since $V_n \leqq V_m$, the number of remaining seats will then be smaller than in the present form of HB.

In order to describe the first property of RE, we follow THEIL ([7], p. 521), who discusses the distance no. 7. Note that

$$\sum_{i=1}^{n} s_i \log(s_i/r_i) = S \sum_{i=1}^{n} q_i \log(q_i/p_i) = SI(\boldsymbol{q}; \boldsymbol{p}),$$

where $p_i = r_i/S$ and $q_i = s_i/S$ ($i = 1, 2, \ldots, n$) define two discrete probability distributions with n possible outcomes (with $\Sigma p_i = \Sigma q_i = 1$). The concept $I(\boldsymbol{q}; \boldsymbol{p})$ is the KULLBACK-LEIBLER information distance between these two probability distributions. It is well-known that (1) $I(\boldsymbol{q}; \boldsymbol{p}) > 0$ if $\boldsymbol{q} \neq \boldsymbol{p}$ and (2) $I(\boldsymbol{p}; \boldsymbol{p}) = 0$. Note that $I(\boldsymbol{q}; \boldsymbol{p})$ takes larger and larger values when $\boldsymbol{q}$ and $\boldsymbol{p}$ are coordinatewise more different. It is interesting to see what happens when $\boldsymbol{q}$ and $\boldsymbol{p}$ are close to each other. Defining $\varepsilon_i := q_i/p_i - 1$, we have, assuming $|\varepsilon_i| < 1$ and using $\Sigma \varepsilon_i p_i = 0$:

$$I(\boldsymbol{q}; \boldsymbol{p}) = \sum_{i=1}^{n} q_i \log(q_i/p_i) = \sum_{i=1}^{n} p_i(1+\varepsilon_i)\log(1+\varepsilon_i) =$$

$$= \sum_{i=1}^{n} p_i(\varepsilon_i + \tfrac{1}{2}\varepsilon_i^2 - 2^{-1}3^{-1}\varepsilon_i^3 + 3^{-1}4^{-1}\varepsilon_i^4 - \ldots) = \tfrac{1}{2}\sum_{i=1}^{n} p_i\varepsilon_i^2 + O(\varepsilon_i^3).$$

Hence, we have $I(\boldsymbol{q}; \boldsymbol{p}) \approx \frac{1}{2}\Sigma(p_i - q_i)^2/p_i$. The right hand side of this approximation is proportional to a chi-square with the p_i's as theoretical probabilities and the q_i's as observed frequencies. Now $I(\boldsymbol{q}; \boldsymbol{p})$ can be interpreted as a measure for the degree to which we are surprised when we are informed that the prior probabilities $p_1, p_2,$

$\ldots, p_n$ are replaced by the posterior probabilities $q_1, q_2, \ldots, q_n$. Interpret now the p_i's again as the proportions of the votes cast on the various parties and the q_i's as the corresponding proportions of the parliamentary seats. Obviously, the electorate ought to be surprised as little as possible. So far THEIL, with some explaining remarks.

Now replacing $\boldsymbol{p}$ and $\boldsymbol{q}$ by $\boldsymbol{r}$ and $\boldsymbol{s}$, respectively, we find, assuming that $|\varepsilon_i| = |q_i/p_i - 1| = |s_i/r_i - 1| < 1$, i.e., $0 < s_i < 2r_i$,

$$\sum_{i=1}^{n} s_i \log(s_i/r_i) = SI(\boldsymbol{q}\,;\boldsymbol{p}) \approx \tfrac{1}{2}S \sum_{i=1}^{n} (p_i - q_i)^2/p_i = \tfrac{1}{2} \sum_{i=1}^{n} (s_i - r_i)^2/r_i.$$

Hence, by using lemma 5.1c, we arrive at the first property of the procedure RE.

Property 1

The procedure RE *"in first approximation" minimizes the "degree to which we are surprised" when we are informed that a given theoretical seat distribution is replaced by the actual one.*

Remarks

Instead of working with $I(\boldsymbol{q}\,;\boldsymbol{p})$, one could equally well work with $I(\boldsymbol{p}\,;\boldsymbol{q})$. This leads to distance no. 6 as a first approximation of the distance function $2\ \Sigma r_i \log(r_i/s_i)$. It may be interesting to notice that the procedure d_6 is, in fact, the so-called method of Equal Proportions (EP) of HUNTINGTON ([2], p. 109), which is presently used in the USA for determining the numbers of representatives of the States in the House of Representatives, proportional to the respective numbers of inhabitants of the States. As to the way of assigning the seats, HUNTINGTON proposes an algorithm which assigns the seats one after the other, starting with $s_i = 1$ for every State. Our algorithm starts, in step 1, with s_i near to r_i, which is more efficient, especially for a House of 435 seats!

As to the second property of RE, we recall that HB favours the greater parties, whereas the procedure SD favours the smaller parties (cf. Section 2). Now using lemma 5.1, and comparing the distances nos. 3, 4 and 5 in table 4, we arrive at the second property.

Property 2

The procedure RE is impartial in the sense that its distance (no. 5) *is situated in between the distance which favours the greater parties* (no. 3) *and that favouring the smaller parties* (no. 4).

The third property was already discussed in Section 4.

Property 3

The procedure RE *produces its solution of the* PR*-problem in a simple and efficient way.*

Competitors with respect to this third property are the procedures d_1 and d_2, for which step 1 of our algorithm delivers, just as for RE, $s_i = \text{round}(r_i)$, so that very few corrections with step 2 or step 3 are needed (provided that $\Sigma r_i = S$). We want to add a few remarks on these procedures d_1 and d_2.

Apparently, the procedures d_1 and d_2 comprise a whole class of procedures for solving the PR-problem, because of the presence of the parameter a (cf. table 4). However, one can easily prove, by using the convexity of the functions $f_i(x) = |x - r_i|^a$ for any real $a \geqq 1$, that if s minimizes $\Sigma|s_i - r_i|^a$ for $a = a_1$, then it also minimizes this distance for $a = a_2$ ($\neq a_1$). The same statement holds for distance no. 2.

Secondly, if the theoretical seat distribution r satisfies $S - n \leqq \sum_{i=1}^{n} [r_i] \leqq S$ (which includes, of course, the one for which $\sum_{i=1}^{n} r_i = S$), then the procedure d_1 is, in fact, equivalent with the method of the Greatest Remainders, also known as the method of ROGET (cf. [3]), which was proposed already in 1792 by HAMILTON (cf. [1]). Its algorithm is as follows. First put $s_i := [r_i]$, and order the fractional parts $\delta_i = r_i - s_i$ in a priority list $\delta_{i_1} \geqq \delta_{i_2} \geqq \ldots \geqq \delta_{i_n}$. Next, give an additional seat to each of the first $S - \Sigma[r_i]$ parties on the list. In Appendix D we show that the method of the greatest remainders and d_1 are equivalent.

A disadvantage of the procedure d_1 (and also of d_2) is demonstrated by the so-called "ALABAMA-PARADOX" (cf. [2]). Given the theoretical seat distribution r and the corresponding actual distribution s, found with d_1. Let it be decided, for some reason, to increase the total number of representatives with 1 (so that also the components of r increase). Then it is possible that a party *loses* one seat as compared with the old situation. An example is given in table 6. Party 3 loses one seat when the total number of representatives is brought from 100 to 101. Procedures $d_3, \ldots, d_7$ are *not* subject to the Alabama-Paradox. This follows from combining the fact that, for the distance d of each of these procedures, we have $\sigma_S(d, r) = \sigma_S(d, cr)$, for any fixed positive constant c (cf. our remarks in Section 4 before lemma 4.1), and lemma 4.1.

Table 6. The Alabama-Paradox for the procedure d_1

i	r_i	s_i	r_i*	s_i
1	45.29	45	45.74	46
2	44.20	44	44.64	45
3	10.51	11	10.62	10
	100.00	100	101.00	101

* The fourth column is 1.01 times the second

6 The weighted vote

In the light of the first four lines of HUNTINGTON's paper [2]:

> "*In the absence of any provision for fractional representation in Congress, the constitutional requirement that the number of representatives of each state shall be proportional to the population of that state cannot be carried out exactly*",

we quote Theorem VIII of the author's doctor's thesis (1976):

> *Consider a house chosen by democratic elections such that on the basis of proportional representation S seats* ($S \geqslant 1$) *are allocated to n parties "in a way as fair as possible", in the proportion* $s_1 : s_2 : \ldots : s_n$ ($s_i \in \mathbb{N}$, $\Sigma s_i = S$). *Suppose that according to the results of the elections the seats should have been divided in the proportion*

$r_1 : r_2 : \ldots : r_n$ $(r_i \in \mathbb{Q}, \Sigma r_i = S)$. *Then in case of votings in the house it should be strongly recommended to give the voting weight* r_i/s_i *to the votes of any representative of party i, instead of the usual weight* 1.

The great advantage of the weighted vote will be clear: the representation of the votes of the electorate in the house is perfect, as it should be, and independent of the procedure used for the determination of the actual seat distribution. In systems with a voting-threshold there are two possibilities to define voting weights. It can either be based only on the votes of the parties admitted to Parliament, or it can be based on *all* the votes recorded by the electorate. We prefer the first possibility.

Table 7. Voting weights for the parties in the Second Chamber

party i	number of votes v_i	number of representatives (HB) s_i	first definition r_i	first definition voting weight r_i/s_i	second definition r_i	second definition voting weight r_i/s_i
1	2,813,793	53	51.7084	0.9756	50.7250	0.9571
2	2,655,391	49	48.7974	0.9959	47.8695	0.9769
3	1,492,689	28	27.4308	0.9797	26.9091	0.9610
4	452,423	8	8.3141	1.0393	8.1560	1.0195
5	177,010	3	3.2529	1.0843	3.1910	1.0637
6	143,481	2	2.6367	1.3184	2.5866	1.2933
7	140,910	3	2.5895	0.8632	2.5402	0.8467
8	79,421	1	1.4595	1.4595	1.4317	1.4317
9	77,972	1	1.4329	1.4329	1.4056	1.4056
10	69,914	1	1.2848	1.2848	1.2604	1.2604
11	59,487	1	1.0932	1.0932	1.0724	1.0724
	8,162,491	150	150.0002		147.1475	
others	158,234					
total	8,320,725					

As an illustration table 7 gives the voting weights, according to both definitions, for the parties in the Second Chamber with the present seat distribution. We derive from it an example which illustrates the importance of this correction. Assume that for approval of a certain proposal a majority of $\frac{2}{3}$ of the votes is required. Now let all representatives vote for the proposal, except those of parties 2 and 8. As a result the proposal is accepted with a majority of 100 votes against 50. However, using the voting weights according to the first definition, the number of votes for approval amounts to 99.7433, so that the proposal is to be rejected.

The voting weight should be (and can easily be) realized in practice as an automatic system: Each representative in the Chamber will have two buttons at his disposal, one for voting for and one for voting against. The votes recorded by the representatives are summed automatically by the system, each with its proper weight. The result of the voting is displayed at the Chairman's desk almost immediately after the voting. Only after new elections (once in every four years, as a rule) the weights will have to be changed in this automatic system.

Appendix

A. *Proof of lemma 4.1* (this proof is a generalization of a similar proof of THEIL and SCHRAGE in [8])

We prove the existence of $s^* \in \sigma_{S+1}(d, r)$ such that $s_i^* \geqq s_i$ $(i = 1, 2, \ldots, n)$. This implies the existence of ${}^*s \in \sigma_{S-1}(d, r)$ with ${}^*s_i \leqq s_i$ $(i = 1, 2, \ldots, n)$. The functions $f_i(.)$ are convex, so

$$f_i(x) - f_i(x-1) \text{ is a non-decreasing function of } x \; (i = 1, 2, \ldots, n) \tag{A.1}$$

For given S, let $s = (s_1, s_2, \ldots, s_n) \in \sigma_S(d, r)$. We define the left and right differences

$$L_i(S) = f_i(s_i - 1) - f_i(s_i) \tag{A.2}$$

and

$$R_i(S) = f_i(s_i) - f_i(s_i + 1) \tag{A.3}$$

where the insertion of S in parentheses after L_i and R_i indicates that s_i, and hence also $f_i(s_i - 1)$, $f_i(s_i)$ and $f_i(s_i + 1)$ depend on S. Suppose that we raise s_i by 1 and lower s_j by 1 $(i \neq j)$, so that $d(r, s)$ increases by

$$f_i(s_i + 1) - f_i(s_i) + f_j(s_j - 1) - f_j(s_j) = L_j(S) - R_i(S).$$

This must be non-negative, because otherwise s would not belong to $\sigma_S(d, r)$. Therefore

$$L_j(S) \geqslant R_i(S), \quad \text{for} \quad i \neq j. \tag{A.4}$$

Next we raise S by 1 (but we don't change r). At least one of the values of $s_1, s_2, \ldots, s_n$ must then increase: let this be s_i, which becomes $s_i + t$, where $t \in \mathbb{N}$, so that (A.2) yields

$$L_i(S+1) = f_i(s_i + t - 1) - f_i(s_i + t). \tag{A.5}$$

The right hand side equals $R_i(S)$ given in (A.3) if $t = 1$, and it is smaller than $R_i(S)$ if $t > 1$, in view of the convexity property (A.1). Therefore,

$$L_i(S+1) \leqslant R_i(S). \tag{A.6}$$

Suppose now that the increase of S by 1 lowers s_j; *we prove that this possibility can be ignored.* If s_j becomes $s_j - u$, where $u \in \mathbb{N}$, (A.3) yields

$$R_j(S+1) = f_j(s_j - u) - f_j(s_j + 1 - u) \tag{A.7}$$

which equals $L_j(S)$ if $u = 1$ and exceeds $L_j(S)$ if $u > 1$, in view of (A.1). Therefore

$$R_j(S+1) \geqslant L_j(S). \tag{A.8}$$

Combining (A.8), (A.4) and (A.6) yields

$$R_j(S+1) \geqslant L_j(S) \geqslant R_i(S) \geqslant L_i(S+1)$$

hence $R_j(S+1) \geqslant L_i(S+1)$. It follows from (A.7) and (A.5) that this can be written as

$$f_i(s_i+t)+f_j(s_j-u) \geqslant f_i(s_i+t-1)+f_j(s_j-u+1).$$

This means that we can lower s_i+t by 1 and raise s_j-u by 1 without increasing $d(\boldsymbol{r}, \boldsymbol{s})$. When we go from S to $S+1$ the number of increases in the components of $\boldsymbol{s}$ must exceed the number of decreases by 1. So, whenever we find a party j with fewer seats under $S+1$ than under S, we can always find a party i with more seats under $S+1$ than under S and transfer one seat from i to j without increasing the distance. These transfers, none of which raises the distance, can be repeated until each party has at least as many seats under $S+1$ as it had under S. This completes the proof.

B. *Proof of lemma 5.1*a

Let n parties be admitted to Parliament, i.e., n parties have scored at least the quota Q_m. For the theoretical seat distribution $\boldsymbol{r}$ we have $r_i = v_i S/V_m$, where V_m is the total numbers of votes recorded by the electorate, including those of the parties not admitted to Parliament. Hence $\sum_{i=1}^n v_i \leqq V_m$ and $\sum_{i=1}^n r_i \leqq S$. Now step 1 of the algorithm for distance no. 3 delivers $s_i = [r_i]$, which agrees with the numbers of seats, initially allocated by the procedure HB. Then we have $\Sigma s_i \leqq \Sigma r_i \leqq S$. Now the procedure HB allocates the remaining seats, one after the other, to the party which shows, *after* the allocation, the greatest number of votes per seat, i.e., the greatest value of $v_i/(s_i+1)$. Step 2 of our algorithm for distance no. 3 allocates the remaining seats, one after the other, to the party for which

$$f_i(s_i+1)-f_i(s_i) = (s_i+1-r_i+\tfrac{1}{2})^2/r_i-(s_i-r_i+\tfrac{1}{2})^2/r_i = 2(-1+(s_i+1)/r_i)$$

is minimal, in other words, for which $r_i/(s_i+1)$ is maximal. This completes the proof, since $r_i = v_i S/V_m$, and S/V_m is a constant, independent of i.

C. *Proof of lemma 5.1*c

For the theoretical seat distribution $\boldsymbol{r}$ we have $r_i = v_i S/V_n$, so that $\sum_{i=1}^n r_i = S$. Step iii of the procedure RE (see Section 3) is obviously the same as step 1 of our algorithm for distance no. 5: $s_i := \text{round}\ (r_i)$, for $i = 1, 2, \ldots, n$. If $\Sigma s_i = S$, we are ready. We show that for $\Sigma s_i < S$ RE and d_5 deliver the same seat distribution. The case $\Sigma s_i > S$ is handled analogously.

If $\Sigma s_i = \Sigma\ \text{round}\ (r_i) < S$, RE performs step iv. This means that a constant $c > 1$ is sought such that $\Sigma\ \text{round}\ (cr_i) = S$. This constant can be found as follows: start with $c = 1$ and let c continuously increase. Then the fractional parts of the numbers cr_i will also increase continuously (passing 1 means starting again at 0, of course), and we stop the first time one of the fractional parts reaches $\frac{1}{2}$. Then we have $cr_i = s_i + \frac{1}{2}$, for some $i = i_0$, and, since c is minimal,

$$c = (s_{i_0} + \tfrac{1}{2})/r_{i_0} = \min_i ((s_i + \tfrac{1}{2})/r_i).$$

We put $s_{i_0} := s_{i_0} + 1$ and the process of increasing c continuously is resumed until $\Sigma s_i = S$ is reached. We can formulate this process as follows: the procedure RE allocates the remaining seats, one after the other, to the party which shows the smallest value of $(s_i + \frac{1}{2})/r_i$ (or the largest value of $v_i/(s_i + \frac{1}{2})$). Step 2 of our algorithm for distance no. 5 allocates the remaining seats, one after the other, to the party for which

$$f_i(s_i+1) - f_i(s_i) = (s_i + 1 - r_i)^2/r_i - (s_i - r_i)^2/r_i = 2(-1 + (s_i + \tfrac{1}{2})/r_i)$$

is minimal. This condition clearly coincides with that of RE, which completes the proof.

D. *Proof that the method of the greatest remainders is equivalent with minimizing* $\Sigma|s_i - r_i|^a$, $a \geqq 1$, *where* r *satisfies* $S - n \leqq \Sigma[r_i] \leqq S$.

Since the value of a ($\geqq 1$) does not matter, we choose $a = 1$. Step 1 of our algorithm yields $s_i = \text{round}\,(r_i)$. If $\Sigma s_i = S$ then we are finished. Assume $\Sigma s_i < S$. The other case is proved analogously. In order to perform step 2, we compare $f_i(s_i+1) - f_i(s_i)$ with $f_j(s_j+1) - f_j(s_j)$, assuming that, in step 1, r_i was rounded off *upwards* and r_j *downwards*. So $s_i = r_i + \gamma_i$, $0 < \gamma_i \leqq \frac{1}{2}$, and $s_j = r_j - \delta_j$, $0 \leqq \delta_j < \frac{1}{2}$. Then we have

$$f_i(s_i+1) - f_i(s_i) = |s_i + 1 - r_i| - |s_i - r_i| = |\gamma_i + 1| - |\gamma_i| = 1$$

and

$$f_j(s_j+1) - f_j(s_j) = |1 - \delta_j| - |-\delta_j| = 1 - 2\delta_j \leqslant 1.$$

Hence, only parties for which r_i was rounded off *downwards* in step 1 need to be considered in step 2. Now $1 - 2\delta_j$ is minimal when δ_j is maximal. So the party with largest fractional part $< \frac{1}{2}$ gets the first seat in step 2, the party with the second largest fractional part the second seat, and so on, until $\Sigma s_i = S$ is reached (that indeed $\Sigma s_i = S$ is reached, follows from the assumption $S - n \leqq \Sigma[r_i]$). This process is obviously the same as the process of, firstly, allocating $[r_i]$ seats to party i, ordering the fractional parts $r_i - [r_i]$, and, secondly, allocating the remaining seats to the parties with the largest fractional parts. This completes the proof.

Acknowledgements

An early draft of this paper was written jointly by Dr. J. H. C. LISMAN and the present author. However, as the amount of mathematics involved grew steadily, at a given stage Dr. LISMAN felt that he could be no longer "co-responsible" for this paper. He modestly withdrew as a co-author, in spite of the persistent objections of the present author. Therefore, I am much indebted to Dr. LISMAN, not only for his share in this research, but also for his stimulating and, in spite of his respectful age, youthful enthousiasm.

I am very grateful to Prof. SCHAAFSMA for many improvements, especially with respect to the presentation of this paper. I also thank the referees for a number of improving remarks.

References

[1] BALINSKI, M. L. and H. P. YOUNG (1977), On Huntington methods of apportionment, SIAM J. Appl. Math., Vol. 33, pp. 607–618.

[2] HUNTINGTON, E. V. (1928), The apportionment of representatives in Congress, Trans. Amer. Math. Soc., Vol. 30, pp. 85–110.

[3] LISMAN, F. J. (1973), Waar zetelt uw stem? Intermediair, No. 25, Reacties en commentaar, ibidem, No. 48.

[4] LISMAN, F. J. and J. H. C. LISMAN (1977), De nieuwe zetelverdeling in de Tweede Kamer, Intermediair, No. 23.

[5] LISMAN, F. J., J. H. C. LISMAN and H. J. J. TE RIELE (1978), De zetelverdeling in de Tweede Kamer, Praktische politicometrie, Economisch-Statistische Berichten, No. 3151, pp. 413–418.

[6] REINDERS, A. (1973), Het Zweedse kiesstelsel, Intermediair, No. 37.

[7] THEIL, H. (1969), The desired political entropy, The Amer. Political Science Rev., Vol. 63, pp. 521–525.

[8] THEIL, H. and L. SCHRAGE (1977), The apportionment problem and the European Parliament, European Economic Rev., Vol. 9, pp. 247–263.

[9] WILLCOX, W. F. (1916), The apportionment of representatives, Amer. Economic Rev., Vol. 6, No. 1, Supplement (March), pp. 1–16.