The Mathematics of Proportional Representation

Joyce Aitchison CMathFIMA

1 Introduction

In the second week of June 2004 some 350 million people in 25 countries had the opportunity to vote for the 732 members of the European Parliament. The elections were run under a system of proportional representation (PR), where each constituency is represented by several MEPs, although the precise formula used for the allocation of seats varies from country to country.

Many countries in mainland Europe have used PR for their national elections for some time, but it is a relatively new idea in the UK. Here we have traditionally used an arrangement whereby each constituency is represented by a single MP who is elected on a first-past-the post system. This has the advantage that each constituency has a delegate in parliament who was elected by name. It has the disadvantage that a party may win a sizeable fraction of the national vote but have few or even no members of parliament. In the UK, elections to the national parliament in Westminster are still conducted on a first-past-the-post system. Most local elections to county, metropolitan or district councils are also conducted in a similar way, although there may be variations when there are two seats available in a ward, each voter has two votes and first-past-the-post becomes first- and second-past-the-post.

In Northern Ireland a single transferable vote is now used for elections to the Northern Ireland assembly and local councils. This system was also used for the Northern Ireland constituency in the European elections. The further devolution of government has seen the introduction of other forms of PR voting into the rest of the UK. In particular, the members of the Scottish Parliament, the National Assembly for Wales and the London Assembly are elected using the additional member system. The first occasion when PR was used across the whole of the UK was in 1999 for the previous elections to the European Parliament.

Details of a great variety of voting systems are given on the webpages of the Electoral Reform Society [1]. A detailed review of the use of PR voting in the UK is given in the report of the Independent Commission on PR [2].

This article concentrates on the mechanics of allocating the seats once the votes have been cast. In particular we consider elections which operate using a single non-transferable vote and a party list, and take as our examples the British elections to the European Parliament (excluding Northern Ireland) and the elections to the London Assembly. Both of these elections took place on 10 June 2004.

2 Elections using party lists

In an election using party lists, the voter casts a single vote for a named party, and each party is allocated a number of seats according to its share of the vote. These seats are taken by individuals from the party's list: in an open list system, the voters have some influence over which individuals are chosen; in a closed list system, the party publishes in advance an ordered list of its individual candidates. If the party gains one seat then this goes to the first person on the list, two seats go the first and second people on the list, and similarly for higher numbers. The different algorithms assign the seats in some order, but these do not confer any ranking once the process is over. A candidate on a list is either elected to a seat or is not.

Note that the term "party" can include an individual independent candidate whose list contains a single name. At the other extreme, there is clearly no point in having more candidates in a party list than there are seats available for the constituency. This is illustrated by the South East of England constituency for the 2004 European elections. This is a relatively large constituency with 10 seats available. There were 13 parties standing for election. Of these, nine fielded a full list of 10 candidates, one fielded 9 candidates, one fielded 8 candidates, one fielded 3 candidates and the final "party" was a single independent candidate.

To analyse how the seats are allocated, consider the election in a single constituency with N seats, and assume that there are M parties standing for election. A PR electoral system tends to attract minority parties and independent candidates to stand for election and so normally $M > N$. Let the parties be labelled $P_1, \ldots, P_M$, and let υ_i be the number of votes cast for P_i. If T is the total number of votes cast then

$$T = \sum_{j=1}^{M} \upsilon_j.$$

Let θ_i be the fraction of votes cast for P_i then

$$\theta_i = \frac{\upsilon_i}{T} \quad \text{and} \quad \sum_{j=1}^{M} \theta_i = 1.$$

In any list-based PR election the aim is to allocate the seats roughly in proportion to the votes cast. Let s_i be the number of seats which will be allocated to P_i. Then in an ideal world s_i would be equal to $N\theta_i$. However the numbers of seats must be integers and so we need to find a way of handling the fractions. Put another way, we need an algorithm for sharing out the seats to give the fairest reflection of the votes cast.

Let the parties be numbered in decreasing order of share of the votes so that

$$\theta_1 \geq \theta_2 \geq \ldots \geq \theta_M.$$

Any fair system will yield

$$s_1 \geq s_2 \geq \ldots \geq s_M,$$

and if $M > N$ then

$$s_1 \geq s_2 \geq \ldots \geq s_N \quad \text{and} \quad s_{N+1} = s_{N+2} = \ldots = s_M = 0.$$

Define the quota of seats due to P_i as q_i where $q_i = N\theta_i$ and let $q_i^{(0)}$ be the integer part of q_i. Since

$$\sum_{j=1}^{M} q_j = \sum_{j=1}^{M} N\theta_j = N$$

it follows that

$$\sum_{j=1}^{M} q_j^{(0)} \le N$$

with equality being very unlikely. Therefore we can make an initial allocation of seats where $s_i^{(0)} = q_i^{(0)}$. In other words we allocate each party a number of seats equal to its quota of seats but round all the fractions down. The remaining seats must now be allocated. There are two different classes of methods for doing this. One class looks at the remainders after rounding and tries to reduce them. The other class (known as divisor methods) scales the quotas, so that when the scaled quotas are rounded down, their sum is exactly equal to N.

2.1 Greatest remainder method

Remainder methods are easier to explain and understand than divisor methods and, as an example, we describe the greatest remainder (GR) method. This is the most widely used form of remainder method and is operated with variations in Italy, Israel and South Africa.

Using the previous notation let party P_i deserve a quota of seats q_i but actually receive s_i. The remainder (or shortfall) is defined as $r_i = q_i - s_i$. Then the greatest remainder method chooses $s_1, s_2, \ldots, s_M$ so that

$$\max_{1\le j\le M}(q_j - s_j) \text{ is minimised subject to } s_i \ge q_i^{(0)}, i = 1, \ldots, M.$$

This is implemented in a simple two-stage process. Firstly we make an initial allocation of $s_i^{(0)} = q_i^{(0)}, i = 1, \ldots, M$. Then the party with the greatest remainder is identified, P_k, and allocated an additional seat. This is equivalent to rounding the fractional part of q_k up rather than down. The party with the next greatest remainder is now allocated an additional seat, and so on until all N seats have been allocated.

Since the remainder for each party after the first stage is less than one and there are M parties, it follows that there are at most M seats to be allocated in the second stage. Provided that no party has already used all the candidates on its list, then each party can gain at most one seat in the second stage. Thus, using this algorithm we have achieved the simple goal of having $s_i \approx q_i$ with the fractional part of q_i rounded either up or down. The final allocation of seats satisfies the simple bounds

$$q_i^{(0)} \le s_i \le q_i^{(0)} + 1.$$

In the above analysis we have defined q_i and $q_i^{(0)}$ in terms of the fraction of votes θ_i. Election results are announced and officially displayed in terms of actual votes, and so it can be more helpful to think in terms of the number of votes needed to win one seat. Define the number of votes per seat, V, as $V = T / N$. Then

$$q_i = N\theta_i = N\frac{\upsilon_i}{T} = \frac{\upsilon_i}{T/N} = \frac{\upsilon_i}{V} \quad \text{and} \quad q_i^{(0)} = \text{int}\left(\frac{\upsilon_i}{V}\right).$$

To demonstrate the GR method we consider the votes cast in the North East of England constituency for the 2004 European elections which are shown in table 1. Note that this was not the allocation method used in this election – this calculation is given as a demonstration only. The total votes cast is 780,491 and there are 3 seats to be allocated. Therefore the number of votes per seat is V = 260,164, and the initial allocation is $s_1^{(0)} = 1$, $s_2^{(0)} = s_3^{(0)} = 0$ giving the Labour party one seat.

Table 1: Application of the GR algorithm - North East England

Party	υ_i	$s_i^{(0)}$	r_i	s_i
Lab	266,057	1	5,893	1
Con	144,969		$\boxed{144,969}$	1
Lib Dem	138,791		$\boxed{138,791}$	1
UKIP	94,887		94,887	
BNP	50,249		50,249	
Ind	39,658		39,658	
Green	37,247		37,247	
Respect	8,663		8,663	

Looking at the remainder column, we see that the greatest remainder is r_2 and so we put $s_2 = s_2^{(0)} + 1 = 1$, giving the Conservatives one seat. Finally the next greatest remainder is r_3 and so we put $s_3 = s_3^{(0)} + 1 = 1$, giving the Liberal Democrats the last seat.

Remainder methods have not been used in elections in the UK. Although the initial allocation is done on the basis of proportionality, the allocation of the remaining seats is done using the actual size of the remainder rather than any relative measure. This gives a bias away from proportionality in favour of the smaller parties for those seats. Whether this is a good thing or not is a matter of political not mathematical judgement.

2.2 The d'Hondt method

The algorithm which was actually used in the European elections in Great Britain is the d'Hondt method, which is the most widely used divisor method. It is named after its inventor Victor d'Hondt (1841–1901), who was a Belgian lawyer and mathematician. The method, which is also known as the highest average method, was first described in 1878.

It is usually described as an algorithm in which there is no initial allocation and all the seats are assigned sequentially. The underlying idea is that a party's vote is divided by a number (the divisor) which increases as the party wins more seats. Thus the party's "vote" in succeeding rounds decreases, allowing parties with lower initial votes to win seats. The divisor in the first round is one (having no effect) and subsequently it is the number of seats gained so far plus one.

To illustrate the method we look at the constituency of Wales for the 2004 European elections, which has N = 4 seats and M = 10 parties. The number of votes won by each party are shown in the first column of table 2. Labour has the highest total, 297,810, and so wins the first round. The Labour figure is now divided by 2 (1 seat + 1) to give a new figure of 148,905. To allocate the second seat we see that the highest total for round two is 177,771 for the Conservatives, who win the seat. Their figure is reduced to 88,886 for the next round. The highest total for the third round is 159,888 for Plaid Cymru, who win the third seat. Their figure is now reduced to 79,944. Finally round four goes to Labour again as its adjusted figure of 148,905 is the highest in column 4. Hence the final allocation is Labour 2, Conservative 1, and Plaid Cymru 1. If the table continued, the new Labour figure would be 99,270 - the original total divided by 3 (2 seats + 1).

Mathematically this procedure solves the problem of finding a scaling factor f which is as small as possible while allowing

Table 2: Application of the d'Hondt algorithm – Wales

Party	Round 1	Round 2	Round 3	Round 4
Lab	297,810	148,905	148,905	148,905
Con	177,771	177,771	88,886	88,886
PC	159,888	159,888	159,888	79,944
UKIP	96,677	96,677	96,677	96,677
Lib Dem	96,116	96,116	96,116	96,116
Green	32,761	32,761	32,761	32,761
BNP	27,135	27,135	27,135	27,135
Fwd Wales	17,280	17,280	17,280	17,280
CDP	6,821	6,821	6,821	6,821
Respect	5,427	5,427	5,427	5,427

$$s_i = \text{int}(f\,\upsilon_i) \quad i = 1, \ldots, M, \quad \text{and} \quad \sum_{j=1}^{M} s_j = N.$$

The value of f is s_k / υ_k where P_k gains the last seat. The d'Hondt algorithm effectively steps through increasing values of f as the seats are allocated. The value of f after r steps is the reciprocal of the maximum entry in the rth column of table 2.

Rewriting this in terms of quotas, we have calculated a scaling factor $\lambda(= fT/N)$ so that

$$s_i = \text{int}(\lambda q_i) \quad i = 1, \ldots, M, \quad \text{and} \quad \sum_{j=1}^{M} s_j = N.$$

From our earlier analysis we know that an allocation of $s_i = q_i^{(0)} = \text{int}(q_i)$ will not fill all N seats. Therefore λ must be greater than one, and we could shorten the process by making an initial allocation $s_i^{(0)} = q_i^{(0)}$ as in the GR method. This requires the initial divisors to be set up correctly (to reflect the seats already allocated), and might make things just too complicated for television commentators! For the results from Wales in table 2, $\lambda = 1.54$.

Without doing any formal analysis, we can see that the d'Hondt method tends to favour the larger parties (in contrast to the GR method described earlier), because the steadily increasing scaling of the larger quotas means that they move pass integral values more quickly than the smaller quotas. Again any comment on the desirability of this is a matter of political judgement.

2.3 Modifications to the allocation process

There are two situations where any allocation process needs to be modified. The first occurs when a party is allocated a seat but has no more candidates on its list to fill it. This is most unlikely to happen to a formal political party who would normally provide a list that is sufficiently long to surpass even their wildest dreams of electoral success. It might happen to a very popular independent candidate who would be unable to take a second seat if allocated, but even this is unlikely. However the modification is trivial - any consideration of the greatest remainder or highest average is taken only amongst those parties who still have candidates available.

The second modification is imposed in advance of the election in the form of a threshold. Here parties can only be allocated seats provided they achieve a certain minimum fraction of the total votes - typically 5%. Remembering that we are considering PR elections this is only appropriate if there are a large number of seats available in one constituency, and is most frequently used when calculating top-up seats in an additional member system (see the London results below). Again the implementation is trivial - any consideration of the greatest remainder or highest average is taken only amongst those parties whose votes satisfy the threshold.

3 Results from the elections, June 2004

3.1 European Parliament

The electorate of England, Scotland and Wales numbers 43 million people and is represented in the European Parliament by a total of 75 MEPs, elected from 11 constituencies. The number of seats per constituency ranges from 3 to 10. In each constituency there were more parties standing for election than seats available ($M > N$) and the pure d'Hondt method was used to allocate the seats. It is reasonable to assume that the voters were fully aware that they had a single vote and a closed list system, but that they were not influenced by the precise method used to allocate the votes. Therefore it is valid to take the votes cast in June 2004 and compare the results of different allocation algorithms. However it is not valid to recalculate these results using a different voting system such as an open list or a single transferable vote, since the voters could and probably would have voted differently (tactically).

We have already analysed the results from Wales in table 2, but it is interesting to note that the application of the GR algorithm would have given the fourth seat to UKIP, not to the Labour Party. On the other hand the results from the North East of England, which were analysed in table 1, are such that any sensible allocation algorithm will give an identical result. This is primarily because there were only three seats. If there had been another seat the situation would have been different.

A good example of the bias of different methods towards or away from smaller parties is seen in the results from the East of England. This was always likely to be an interesting result

Table 3: Comparison of the d'Hondt and GR algorithms - East of England

Party	Total votes	Seats	
		d'Hondt	GR
Con	465,526	3	2
UKIP	296,160	2	1
Lab	244,929	1	1
Lib Dem	211,378	1	1
Ind - Bell	93,028	0	1
Green	84,068	0	1
BNP	65,557	0	0
Eng Dem	26,807	0	0
Respect	13,904	0	0
Ind-Naisbitt	5,137	0	0
ProLife	3,730	0	0

because of the presence of the well-known independent Martin Bell and the total of seven seats. The results are shown in table 3 where the actual seats allocated using the d'Hondt algorithm are compared with those which would have been obtained using the GR method. We see that the d'Hondt allocation is (over)generous to the largest party giving the Conservatives three seats to Labour's one when the ratio of votes is 2:1. However the GR allocation (over)favours the smaller parties since the Conservatives now get two seats to the Green party's or Bell's one when the ratio of the votes is 5:1. Clearly a fairer allocation would have been the average of the two, but that would have involved fractions!

3.2 London Assembly

The British elections were held on 10 June 2004, and in many areas there were also local elections on that day. The prize for the most complicated election day goes to London, where voters could register a total of five votes. Two of these were for the Mayor of London who was elected on a Supplementary Vote System - voters registering their first and second choices. This is an appropriate system where individual personalities are as important as parties, and is very different from the arrangements described above.

A single vote was used for the London constituency in the elections to the European Parliament as in the rest of Great Britain, and the remaining two votes were used to elect the 25 members of the London Assembly using the Additional Member System. Here one vote is used to elect a constituency representative using first-past-the-post. The second vote is cast for additional members from party lists representing the whole region and the seats are assigned so that the overall result (constituency + additional members) is roughly proportional to the votes cast for parties. There were 14 constituencies (made up of 2 or 3 boroughs) each returning one constituency member, and 11 additional seats which were contested London-wide.

The allocation of the 11 additional seats was done using a modified d'Hondt algorithm with a threshold of 5%. This is the level below which candidates lose their deposit in a British general election and so seems sensible. The constituency seats are allocated first: in this case, Conservative candidates won nine of the constituencies and Labour candidates won five. Using our previous notation we then have $s_1^{(0)} = 9$, $s_2^{(0)} = 5$, $s_3^{(0)} = \ldots = s_M^{(0)} = 0$. The process for allocating the remainder of the seats is shown in table 4. Note that the figures for the first round of the allocation are scaled using the constituency seats already won - eg the Conservative figure is 533,696/(9+1), and those parties with less than 5% of the vote are shown as having zero. The allocation then proceeds and the final number of seats and the calculation which produced the final seat are shown. We see that the Conservatives do not win any additional seats, but note that they cannot lose their constituency seats however poor their showing in the vote for additional members.

It could be argued that the London-wide seats should be filled independently of the allocation of the constituency seats. If we apply the pure d'Hondt algorithm to the London-wide votes and allocate 11 seats from scratch, then the result for the additional seats is Conservatives 4, Labour 3, Lib Dem 2, Green 1, UKIP 1. Obviously this compounds the lead that the Conservative and Labour parties already have from the constituency seats. Note

Table 4: London Assembly - Allocation of London-wide seats

Party	v_i	$s_i^{(0)}$	Round 1	Round 2		Round 11	s_i
Con	533,696	9	53,370	53,370	...	53,370	9
Lab	468,247	5	78,041	78,041	...	58,531	7
Lib Dem	316,218		$\boxed{316,218}$	158,109	...	$\boxed{63,244}$	5
Green	160,445		160,445	$\boxed{160,445}$	...	53,482	2
UKIP	156,780		156,780	156,780	...	52,260	2
BNP	90,365						
Respect	87,533						
CPA	54,914						
ADC	4,968						

that we do not need to apply a threshold since only 11 seats are being allocated.

At the other extreme we could allocate the additional seats within a d'Hondt algorithm as shown in table 4, but without a threshold. In this case both the BNP and Respect would gain a seat despite having less than 5% of the vote. Initially this may seem strange in an allocation of 11 seats, when a minimum nearer 10% would normally be expected to win a seat. However here we are allocating seats 15–25 of a 25 seat assembly, and, unless the individual constituency voting patterns deviate wildly from the London-wide voting, then the final results will not differ greatly from a simple allocation of 25 seats. Hence it could be expected that a seat might be won with less than 5% of the vote, which is why (ironically) the threshold is set.

Finally it is worth noting that the number of spoiled ballot papers in the London elections was higher than normal, but this is hardly surprising given the range of electoral systems which were present at one time.

4 Conclusion

The design and implementation of "fair" electoral systems is both difficult and politically sensitive. However it is helped if there is a clear understanding of the underlying mathematics. Equally important is the explanation of this mathematics to the voting public. Television coverage of elections has not really moved on since the days of two parties, black and white TV and the "swingometer". Although there are now almost instant, sophisticated graphics, the discussion of the results of the European elections did not really explain what was happening. As we have seen the mathematics is not difficult, and so it would be good to see a more informed coverage in future elections.❑

REFERENCES

[1] Electoral Reform Society, http://www.electoral-reform.org.uk/.
[2] Independent Commission on PR, "Changed Voting. Changed Politics. Lessons of Britain's Experience of PR since 1997". The Constitution Unit, School of Public Policy, UCL, London. 2003.

Glossary

ADC	Alliance for Diversity in the Community
BNP	British National Party

CDP	Christian Democratic Party
Con	Conservative
CPA	Christian Peoples Alliance
Eng Dem	English Democrats Party
Fwd Wales	Forward Wales
Green	Green
Ind	Independent
Lab	Labour
Lib Dem	Liberal Democrat
PC	Plaid Cymru
ProLife	ProLife
Respect	Respect - The Unity Coalition
UKIP	UK Independence Party

Joyce Aitchison is now a freelance lecturer and researcher, following a career in higher education. Her research activities are in the mathematical and computational modelling of scientific and industrial problems, with particular interests in partial differential equations and free-surface flows. Although originally trained as a numerical analyst, she is more interested in the solution of problems than in promoting particular methods. She admits to spending every election night watching the television coverage of the results.