Comparing votes and seats with a diagonal (dis-) proportionality measure, using the slope-diagonal deviation (SDD) with cosine, sine and sign

Colignatus, Thomas

Thomas Cool Consultancy Econometrics

17 August 2017

Comparing votes and seats with a diagonal (dis-) proportionality measure, using the slope-diagonal deviation (SDD) with cosine, sine and sign

Thomas Colignatus
August 24 2017

JEL
A100 General Economics: General
D710 Social Choice; Clubs; Committees; Associations,
D720 Political Processes: Rent-seeking, Lobbying, Elections, Legislatures, and Voting Behavior
D630 Equity, Justice, Inequality, and Other Normative Criteria and Measurement

MSC2010
00A69 General applied mathematics
28A75 Measure and integration. Length, area, volume, other geometric measure theory
62J20 Statistics. Diagnotics
97M70 Mathematics education. Behavioral and social sciences

Abstract

Let v be a vector of votes for parties and s a vector of their seats gained in the House of Commons or the House of Representatives - with a single zero for the lumped category of "Other", of the wasted vote for parties that got votes but no seats. Let $V = 1'v$ be total turnout and $S = 1's$ the total number of seats, and $w = v / V$ and $z = s / S$. Then $k = \text{Cos}[w, z]$ is a symmetric measure of similarity of the two vectors, θ = ArcCos[k] is the angle between the two vectors, and Sin[θ] = Sqrt[1 – b p] is a measure of disproportionality along the diagonal in {w, z} space. The geometry that uses Sin appears to be less sensitive than voters, representatives and researchers are to disproportionalities. This likely relates to the Weber-Fechner law. A disproportionality measure with improved sensitivity for human judgement is 10 √Sin[θ]. This puts an emphasis on the first digits of a scale of 10, which can be seen as an inverse (Bart Simpson) report card. The suggested measure has a sound basis in the theory of voting and statistics. The measure of 10 √Sin[θ] satisfies the properties of a metric and may be called the slope-diagonal deviation (SDD) metric. The cosine is the geometric mean of the slope b of the regression through the origin of z given w and slope p of w given z. The sine uses the deviation of this mean from the diagonal. The paper provides (i) theoretical foundations, (ii) evaluation of the relevant literature in voting theory and statistics, (iii) example outcomes of both theoretical cases and the 2017 elections in Holland, France and the UK, and (iv) comparison to other disproportionality measures and scores on criteria. Using criteria that are accepted in the voting literature, SDD appears to be better than currently available measures. It is rather amazing that the measure has not been developed a long time ago and been used for long. My search in the textbooks and literature has its limits however. A confusing aspect of variables in the unit simplex is that "proportionality" concerns only the diagonal in the {w, z} scatter plot while generally (e.g. in non-normalised space) any line through the origin is proportional.

Keywords

General Economics, Social Choice, Social Welfare, Election, Majority Rule, Parliament, Party System, Representation, Proportion, District, Voting, Seat, Metric, Euclid, Distance, Cosine, Sine, Gallagher, Loosemore-Hanby, Sainte-Laguë, Largest Remainder, Webster, Jefferson, Hamilton, Slope Diagonal Deviation, Correlation, Diagonal regression, Regression through the origin, Apportionment, Disproportionality, Equity, Inequality, Lorenz, Gini coefficient

1. Introduction

We consider the situation of a vector *v* of votes, with a total number of votes *V.* People vote for different parties. The length of the vector is the number of parties. Subsequently there is a vector *s* of the seats, with a total number of seats *S.* There are shares or proportions $w = v / V$ and $z = s / S$, normalised to 1 onto the unit simplex, or often 100 percent. Then $q = V / S$ is the natural quota, or number of votes to cover a seat. Voters may vote for parties who get no seats, also called the "wasted vote". The vectors still have the same length, and standardly the zero seats are collected in 1 category "Other".

The literature on voting has a rich discussion about potential measures for the (dis-) proportionality of the allocations of *v* and *s.* Taagepera & Grofman (2003:673) rightly write about a "zoo of indices proposed and used by various researchers", and they score 19 indices on 12 criteria. [1] Overview discussions are by e.g. Taagepera & Shugart (1989), Gallagher (1991), Taagepera & Grofman (2003), Kestelman (2005), Karpov (2008) and Renwick (2015). Malkevitch (2002) gives a discussion in an AMS feature column. Perhaps for historical reasons this literature has focused on the disproportionality format, in which 0 indicates *proportionality* and in which 1 might be *full disproportionality* (if the measure is normalised on [0, 1]). We also adopt this latter format.

To give a taste of the problem, the next section presents graphs on the outcomes of the elections in 2017 for the House of Commons in the UK, the Legislative in France, and the 2nd Chamber of Parliament in Holland. These bodies compare to the US House of Representatives (rather than the Senate or House of Lords).

The subsequent section presents the newly suggested measure so that it is clear what it amounts to. Subsequently we clarify its motivation and reasoning. Subsequently there are some derivative comments, like on other measures that statistics provides. We close with some theoretical examples.

A heuristic is this: the voting literature has the Webster / Sainte-Laguë (WSL) measure $\mathrm{Sum}[w\,(z / w - 1)^2]$, as the weighed sum of the squared deviations of the ratios from the ideal of unity provided by the diagonal. If we would try to make WSL symmetric then this causes division by zero, since parties meet with zero seats (the "wasted vote"). An insight is: the idea to compare each ratio to unity is overdone, because the criterion or proportionality rather applies to the whole situation, and we should not be distracted by single cases. If *b* is the slope of the regression of $z = b\, w$, then $(b - 1)^2$ is a measure on the regression coefficient. This shifts the focus from individual parties to the *relation* between *z* and *w*. Thus we now consider the *slope-diagonal deviation*. This gives scope for symmetry by also looking at the regression $w = p\, z$, so that we can compare *b* and *p.*

A proportional relationship is best described by the line $\lambda\, y + \mu\, x = 0$, which coefficients may be normalised on the unit circle. For nonzero λ this reduces to $y = T\, x$ with $T = - \mu / \lambda$. This need not be the proportionality that is intended for electoral systems. There we look at the $\{w, z\}$ space with $z = b\, w + \mathrm{e}$. Any proportional relation $s = T\, v$ reduces there to diagonal $z = w$ because $1's = T\, 1'v$. Redefining *T* as a target slope in $\{w, z\}$, we have a condition of minimal $(b - T)^2 + (1/p - T)^2$, and a nonzero value would indicate disproportionality with respect to the standard *T* of the field of study considered. PM. There might be a confusion with dispersion. Dispersion implies disproportionality but disproportionality might not need to imply dispersion.

The linear algebra in this paper should not be confused with statistical methods. Various statistical methods use the same linear algebra but statistics also involves assumptions on distributions: and we will make no such assumptions. However, when the linear algebra in this paper results into a new measure, then this measure can be used for new statistics again.

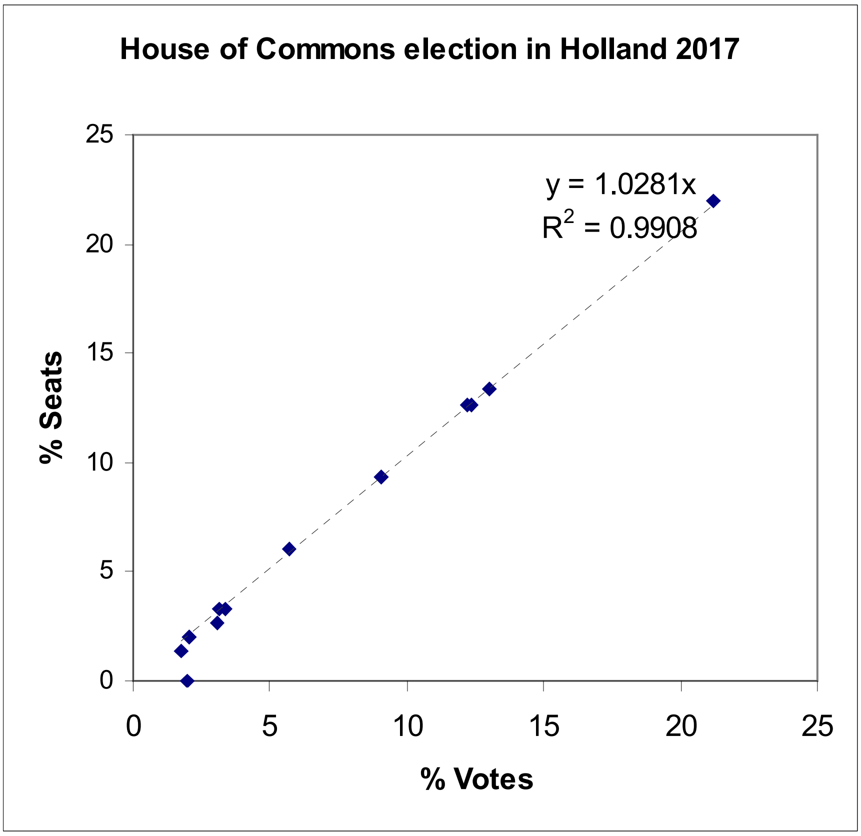
[1] (i) The Taagepera & Grofman (2003) definition of the Gini differs from Colignatus (2017b) which fundamentally affects the major conclusion. (ii) One of their criteria is that there is no explicit need for the number of parties (like in the Chi Square index for the degrees of freedom): this is curious since the measures use *s* and *v* that have the length of the number of parties. (iii) Goldenberg & Fisher (2017) criticise the finding on Dalton's principle of transfer.

2. *Graphs for election results in 2017 in Holland, France and UK*

For the elections in Holland, France and UK in 2017 there is also Colignatus (2017b) for the PR Lorenz curve and PR Gini coefficient on electoral inequality (inequality as a measure is not symmetric as disproportionality must be).

On March 15 2017, Dutch voters elected the parties for the 2nd Chamber of Parliament with 150 seats. Figure 1 gives the regression of $z = s / S$ given $w = v / V$. Holland has Proportional Representation (PR). There is the natural threshold that parties must meet the quota of 1/150 = 0.67% to gain a seat. This low threshold invites competition to the establishment. Yet 2% of the vote or 3 seats are "wasted" on small parties that gain votes but that do not meet the quota. Holland could achieve better PR if it left 2% or 3 of the seats empty. An alternative is to adopt a qualified majority rule. Now 76 seats represent a majority of Floor[S / 2] + 1. Turnout V consists of the vote for parties that got elected Ve and the wasted vote W, including invalid ballots. Thus $V = Ve + W$. Majority threshold $V / 2$ should correspond to $S / 2$. A qualified majority qm on Ve gives the equation $qm\ Ve\ (S / V) = ½\ S$. This gives $qm = ½ / (1 – W / V)$. Thus for a better PR score in Holland , the 2nd chamber of Dutch parliament should not use a majority of 76 but of 77, at least assuming that the wasted vote go either way.

Figure 1



On June 11 and 18, French voters elected the Legislative with 577 seats. Figure 2 gives the regression of z on w. Observe that these regressions do not use a constant, and that the display is around the diagonal (with same scales of horizontal and vertical axes). France has District Representation (DR). The election in districts is in two rounds. We may assume that the first round gives the first preferences of the voters and that the second round involves strategic voting for elimination of the worst option. The horizontal axis uses the votes of the first round and the vertical axis uses seats of the second round.

Figure 2

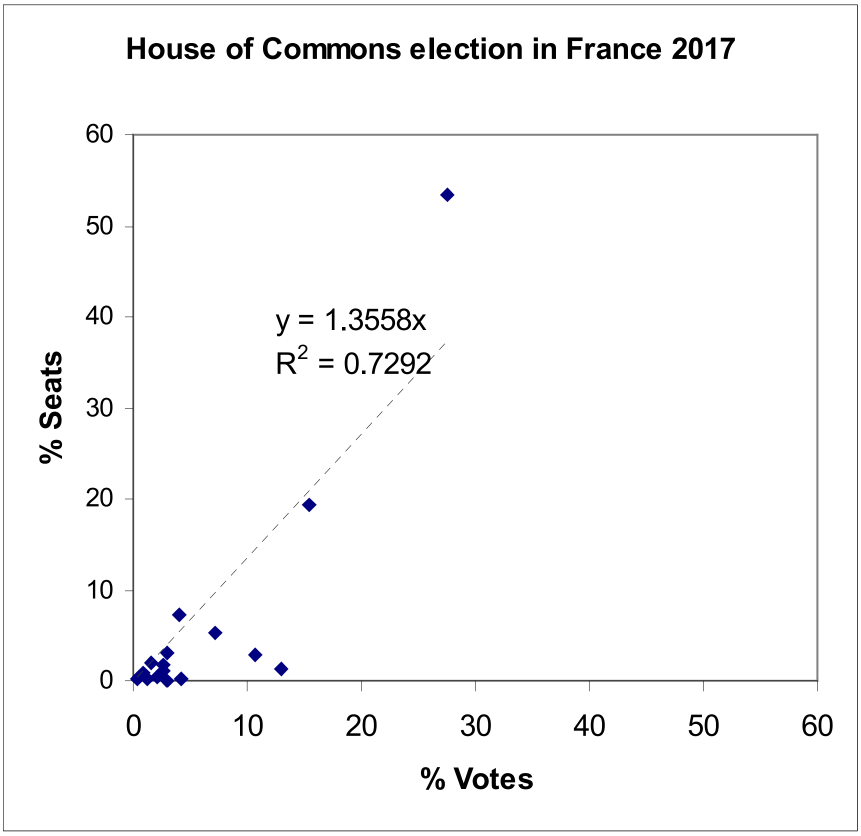
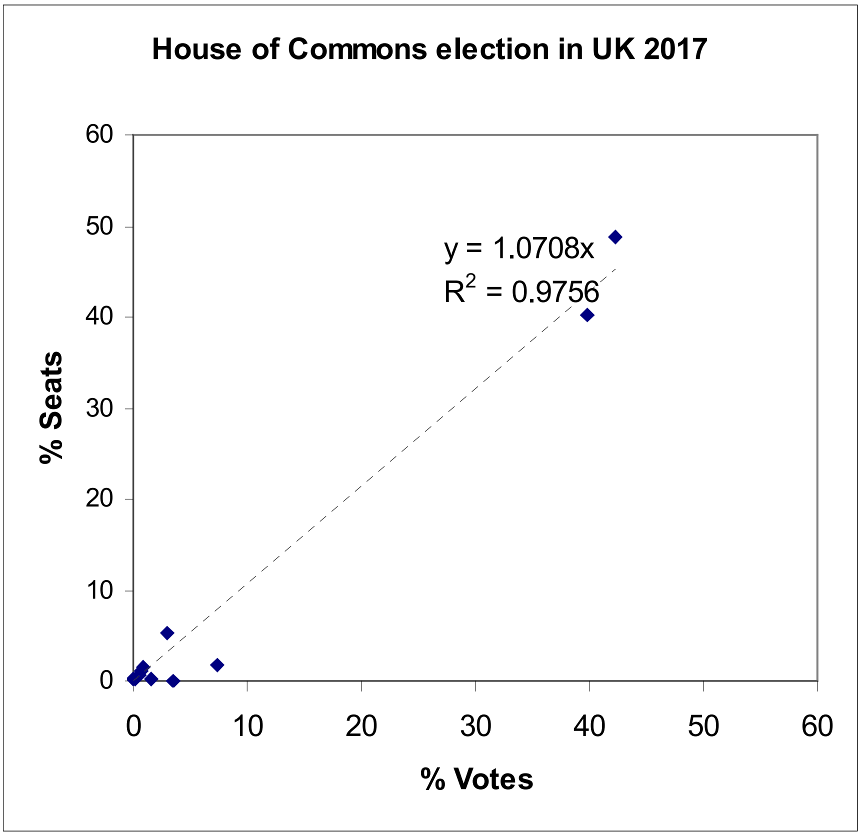


Figure 3



On June 8 2017, voters in the UK elected the House of Commons with 650 seats. Figure 3 gives the regression of *z* on *w* The UK has District Representation (DR), in one round, so that voters may adopt strategic voting to keep out the worst alternative that stands a change of winning. The 2017 UK election seems rather proportional since the voters apparently returned to the model of voting either Conservatives or Labour. However, the outcome can be called "masked" because we do not know the first preferences.

3. Unit proportionality as an approach to power preservation

3.1. The Status Quo and the meaning of the majority rule

It appears necessary here to mention a key aspect that is also highlighted in Colignatus 2014), *Voting Theory for Democracy* (VTFD). This is the role of the Status Quo.

The basic situation in voting has a Status Quo. The issue on the table is that we consider alternatives to the Status Quo. Only those options are relevant that are Pareto Improving, i.e. that some advance while none lose. The Pareto condition thus gives a minority veto rights against being plundered. Infringements like a rail track in the frontyard will require compensations. Commonly there are more Pareto options, whence there is a deadlock in terms of the Pareto rule, that this rule itself cannot resolve. Then majority voting might be used to break this Pareto-deadlock. Many people tend to forget that majority voting is mainly a deadlock breaking rule. A political body that would use a majority rule without such protection of the Status Quo and minority rights would hardly be called a democracy.

When voting for a new House of Commons then it is generally considered no option to leave the seats empty. In this case there would be no status quo. A situation without a status quo tends to be rather exceptional. National elections however have a high profile, and their lack of a status quo might cause people to forget about the role of the status quo for voting situations in general.

3.2. Unit proportionality as a power preserving rule

The use of majorities implies the notion of (coalition-) power. Measures are the indices by Shapley-Shubik ("pivotal") or Penrose-Banzhaf ("critical"). The measures have different values but work out the same, though see Laakso (1980). The approach of ordering is also used for the Shapley value, and thus one tends to choose for the SS score, though PB tends to be mentioned more often in the voting literature perhaps because it originated there.

The key point is to be aware of the mapping from the electorate to representatives:

- The power index would hold for *coalitions amongst the electorate*.
- Proportionality then is a *power preserving rule,* such that the same coalitions from amongst the electorare are also possible amongst the representative body.

Voters, divided up along their party vote at the time of the election, can form various coalitions, and these possible coalitions are preserved in Proportional Representation (PR) by allocating seats to parties in proportion to the vote.

The proportionality will not be perfect:

- There are problems of apportionment, e.g. that some parties may get too few votes to meet the natural quota.
- Within PR there likely will be strategic voting w.r.t. the hoped-for coalition. This is less relevant here, since it would be part of free choice and not something that the system forces. Voters within PR do not have a *necessity* for strategy, like voters within District Representation (DR) do.

Though the property of power preservation is imperfect we may still accept that PR is power preserving in an overall sense. DR has entirely different objectives than PR in this respect.

3.3. Reasons for measuring

We thus arrive at these observations regarding the reasons why we would want to measure disproportionality:

- Proportionality is a power preserving rule, such that majorities *in the votes v* would be preserved *in the apportioned seats s,* and it suffices that we make sure that the proportions are the same.
- Present methods for apportionment are rather crude and focus on proportionality only. It would be a next step in apportionment when parliaments would apportion the seats in such manner that coalition powers are preserved indeed.
- Given that PR and DR have different objectives, the measurement of disproportionality has limited value in comparing these electoral systems. Yet, when the topic arises, it will still be helpful when the measure on disproportionality is sound.

4. Definition of the "slope-diagonal deviation" (SDD) disproportionality measure

4.1. Definition of the SDD measure

Variables x and y are considered to be proportional when $y = b\ x$, thus as a ray through the origin without a constant. Seat allocation is a zero-sum game, and thus we are interested in the line $y = x$ and thus with slope $b = 1$. The suggested "slope-diagonal deviation" (SDD) measure, see Table 1, uses the normalised vectors w and z to define a disproportionality measure around the slope b, in this case of $z = b\ w$, also considering the alternative $w = p\ z$. Some readers may directly understand what the measure does. For others the table is an introduction to be alert to the terms that will be explained in the subsequent sections.

Table 1. Definition of the "slope-diagonal deviation" measure

$w = v / V$ and $z = s / S$	normalisation to 1, or the unit simplex (frequently 100)
$b = z'\ w / w'\ w$	slope of regression through the origin of z given w
$p = z'\ w / z'\ z$	slope of regression through the origin of w given z
k = *cosine*[v, s] = Sqrt[$b\ p$]	the *geometric average slope,* a form of "similarity"
d = Sqrt[$1 - k^2$] = *sine*[v, s]	metric, distance of *slope* k from *slope* 1 Sin = Sqrt[$1 - b\ p$]
disp[d] = $(100\ d)^{1/f}$	transform with sensitivity $f \geq 1$, standard $f = 2$
sign = If[sign[b] < 0, -1, 1]	a zero covariance should not affect the sign
SDD[v, s] = *sign disp*[d]	include the sign for majority switches (and not seeing the measure as a slope itself)

In regression analysis we find both slope and dispersion. The present approach has a focus on the slope, but not at full neglect of dispersion. Both Cos and correlation R are measures of association and slope. The Pearson correlation coefficient R is defined as R = Cos[x – Mean[x], y – Mean[y]]. When y and x are standardized (by subtraction of mean and division by standard deviation) then $y = R\ x$. Thus $R = 1$ fits the notion of proportionality (for such standardized variables). Thus both Sqrt[$1 - R^2$] and Sin = Sqrt[$1 - \text{Cos}^2$] are measures of distance. Subsequently, we don't want to use centered data. A vote of {50, 50} and a seat allocation of {50, 50} is perfectly proportional but causes infinities for R. We still want to keep the regression through the origin. This leads to the definitions in Table 1. See Colignatus (2006) on the sample distribution of the adjusted coefficient of determination.

4.2. Weber-Fechner consideration

Geometry may be less sensitive to disproportionality than voters are. A difference of a seat in the Dutch parliament of 150 (0.67%) or the UK parliament of 650 (0.15%) should not be easily overlooked. Such a statement on sensitivity might seem to be a statement from a Social Welfare Function (SWF) and then seem arbitrary. However, it will more likely relate to the Weber-Fechner law. [2] Disproportionality in the higher ranges is not so relevant but small deviations from proportionality attract our interest.

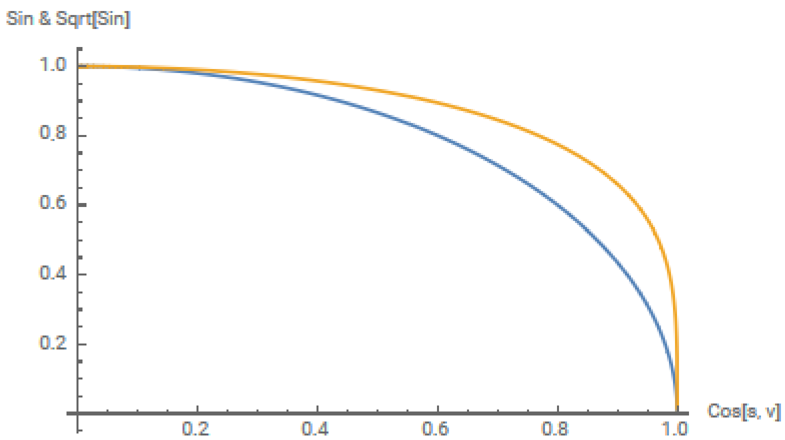
Figure 4 shows how Sin[*v, s*] follows from the estimated Cos[*v, s*]. [3] A sensitivity of $f = 2$ means using √Sin. Looking at some calibrating cases, it appears wise to select the standard value of $f = 2$ indeed. (PM. In Figure 4 the factors 100 and 10 are not included yet.) Factors 10 or 100 can be introduced to eliminate leading zeros. With 100, then $f = 1$ gives the sine values on a scale of 0 to 100. This should not be read as a percentage though. The suggested standard $f = 2$ takes the square root (keeping the sign) and gives values on a scale of 0 to 10. The latter can be read as an inverted (Bart Simpson) report card. Though the SDD index is straightforward math(s), we likely should not attach great value to the second digit. The use of 10 √Sin seems advisable overall.

Table 2 reviews two cases. Case 1 with large disproportionality is modestly scored as 19.6 by the Sin on a scale of 100. The outcome of 19.6 with $f = 1$ may be geometrically straightforward. Yet $f = 2$ better conveys the sizeable disproportionality, with a score of 4.4 on a scale of 10, or 44 or a scale of 100. Case 2 with a small disportionality (that flips the majority) still gets –2 on a scale of 10 while the unadjusted Sin gives –4 on a scale of 100. The latter is unconvincingly insensitive, and thus we may prefer $f = 2$. PM 1. The sign indicates the flipping of majority. PM 2. The differences in scales make a design-phase somewhat more complicated. Once the suggested standard $f = 2$ would be adopted in general, then common comparisons can be on the scale of 10 only, without this potential source of confusion on scale 10 or scale 100.

Table 2. Sensitivity and scale 10 or 100

Votes	Seats	$f = 2$ and *sign* 10 √Sin	$f = 1$ and *sign* 100 Sin
{50, 50}	{60, 40}	4.4	19.6
{51, 49}	{49, 51}	-2	-4

Figure 4. Sensitivity enhancement with $\mathrm{Sin}[v, s]^{1/f} = \sqrt{[1 - \mathrm{Cos}[v, s]^2]}^{\,1/f}$



[2] https://en.wikipedia.org/wiki/Weber%E2%80%93Fechner_law
[3] Created at Wolfram Alpha as Plot[{Sqrt[1 - x^2], Abs[1 - x^2]^(1/4)}, {x, 0, 1}, AxesLabel -> {"Cos[s, v]", "Sin & Sqrt[Sin]"}]

4.3. The properties for a metric are satisfied

Van Dongen & Enright (2012) discuss metrics that involve cosine and Pearson correlation. Their exposition is like taylored for our purposes. [4] 1 – Cos is a distance too but not a metric. Editing their results to our transform gives Table 3. As they did the basic work for us, our own present paper only considers the relevance for the context of electoral studies, and optimises particulars of application. (Though it may perhaps be remarked that I already had the core of this paper before I had the keywords to locate their paper.)

Table 3. Relevant properties of distance and metric

Quadrant I	1 – Cos[*v, s*]	*d* = Sin[*v, s*]	(100 *d*)^(1 / *f*) for *f* ≥ 1
Nonnegative	Yes	Yes	Yes
Zero iff v = s	Yes	Yes	Yes
Symmetric	Yes	Yes	Yes
Triangle inequality	No	Yes	Yes
Thus metric	No	Yes	Yes

Thus the proposed measure Abs[DDS[*v, s*]] is a metric. The negative sign only is an additional feature to help identify majority switches. If we are not interested in this, then we take the absolute value (and we should avoid squaring).

4.4. Scores for Holland, France and UK in 2017

Table 4 reviews the scores for Holland, France and UK in 2017, for the elections discussed in **Section** 2. Some readers will recognise some scores that are traditional in the literature on electoral systems. Other readers would see these scores as terms to be alert on in the following discussion, and then return to this later again.

It is useful to be aware of *which* is *what*. Presently we are in a design phase. The present discussion is targeted on clarifying that the measure in row 1 is the best measure to adopt in the literature on voting and electoral systems, similar to the Richter scale for earthquakes.

Table 4. Scores for Holland, France and UK 2017

	2017	Range	UK	France	Holland	Hol-SDD
	PR or DR		DR	DR 2	PR	PR+
	Masked		Yes	No	No	No
1	10 Sqrt Sin	0 - 10	3.7	6.9	2.5	1.5
2	100 Sqrt Sin	0 - 100	37.1	68.7	25.0	15.5
3	100 Sin	0 - 100	13.8	47.2	6.2	2.4
4	Gini	0 - 100	15.6	41.6	3.6	1.7
5	WSL	0 - inf	12.7	54.8	2.3	0.1
6	Euclid / 1.4	0 - inf	6.8	21.5	1.7	0.6
7	LHI	0 - inf	10.5	33.8	3.0	1.3
8	#Seats		650	577	150	150
9	#Parties		9	16	13	14
10	ENPs		2.5	3.0	8.1	8.7
11	ENPv		2.9	7.1	8.6	8.6
12	Slope (c = 0)		1.071	1.356	1.028	0.997
13	R^2		0.976	0.729	0.991	0.999
	PM. Unweighed regression					

[4] A portal and no reference: https://en.wikipedia.org/wiki/Metric_(mathematics)

Comments are:

(1) While this paper concerns a design-phase, and while we thus should be alert on the distinction between the scale of 10 or 100, I noticed that I myself was confused on this on some occasions. Thus rows 1 and 2 emphasize that row 1 uses scale 10 while row 2 uses the same measure but on scale 100. This means that row 2 simply multiplies row 1 by 10. Normally we would prefer row 1 over row 2 once this confusion is out of the way.
(2) Holland has PR and the Gini (row 4) has a fine low score of 3.6 (or a scale of 100). The Gini includes the wasted vote of 2%. This conforms with the Sin (row 3) of 6.2 (on a scale of 100). However, the plain Gini does not correct for Weber-Fechner. Originally I was pleased that Holland scored well on proportionality but subsequently this discussion puts more emphasis on the wasted vote. The suggested proper measure (row 1) is sensitive to disproportionality with $f = 2$. This generates a staggering 2.5 on a scale of 10 (row 1) (or 25 on row 2). The cause of this disproportionality are the wasted vote to parties who get zero seats. The answer would be that Holland allows empty seats or adopts a qualified majority rule that respects this wasted vote. Holland also uses D'Hondt.
(3) For comparison, a counterfactual case for Holland with PR+ has been created, with the seats relocated as shown in **Appendix F**. The 2% wasted vote are included as empty seats of a 14th party. The scores improve remarkably. Also remarkable is that the SDD measure still has a relatively high value of 1.5 on a scale of 10. This must be the unavoidable result of squeezing 10 million votes into 150 seats, with the number of parties that provide more causes for errors. There is no need to do such calculations for France and UK and produce charts on these, because the outcomes will be similar to the vote shares and dots along the diagonal.
(4) France has a staggering disproportionality of 6.9 (on a scale of 10).
(5) The UK has a disproportionality of 3.7 (on a scale of 10) This is sizable, see also the PR Gini of 15.6%. This meaure is quite uncertain however. The data derive from its system of DR that masks first preferences. We actually do not know what the UK voters actually prefer (also because exit polls apparantly do not ask this question on strategic voting).
(6) My own preference originally was for the Gini (row 4). (i) Colignatus (2017b) explains that parties must be sorted on their s / v scores, to arrive at the proper measurement on the PR Lorenz and PR Gini. However, it is not inconceivable that researchers might make errors on such sorting. Thus a measure is to be preferred that does not depend upon sorting. This leads to SDD (row 1).It is somewhat amazing that considerations on sorting cause such a more fundamental consideration. (ii) However, once SDD had been constructed and then scored on the Taagepera & Grofman (2003) criteria, it was highlighted again that disproportionality is symmetric and that the Gini inequality measure is not symmetric. See the discussion about Table 12. (iii) Colignatus (2017b) doesn't include the Weber-Fechner notion on sensitivity yet. Given this more fundamental suggestion for SDD there is no need to adapt the PR Gini for sensitivity too.
(7) To rephrase: The Gini tends to correlate with the Sin measure, yet not fully because of symmetry. Thus we may adopt row 1, that also includes adequate sensitivity.
(8) Webster / Sainte-laguë (WSL) (row 5) tends to correlate with Gini (row 4) or 100 Sin (row 3). WSL is mentioned by Gallagher (1991) as a likely best standard. See the discussion below however.
(9) Paradoxially, the Gallagher index (labeled for more readers as Euclid / √2) (row 6) has grown to be somewhat of a practical standard in voting literature, contrary to the Gallagher (1991) statement. If we are not uncritical then we may live with this index, that rightly orders Holland < UK < France on a disproportionality scale. However, the proposed index SDD (row 1) better relates to WSL and has more sensitivity to disproportionality.
(10) The indices on the "effective number of parties" are about concentration and not about proportionality.

5. *Different purposes in voting literature and statistics in general*

Vectors s and z have been created by human design upon v, and not by some natural process as in common statistics. Doing a test on s / v commonly does not invoke notions of assumptions on normality in random allocation, for example. The current (inoptimal) disproportionality measures (WSL, Loosemore-Hanby, Gallagher) have frequently been used to compare outcomes of electoral systems across countries, but such comparisons have limited value because of the different designs involved anyway. Taagepera & Grofman (2003) mention also some other reasons for a disproportionality measure: (i) what they call "vote splitting" (better: votes for different bodies): comparison on President, Senate, House, or regional elections, (ii) what they call "volatility" (better: votes for different occasions): comparison on years in similar settings (both votes and seats).

There appears to be some distance between the voting literature on disproportionality and the statistics literature on association, correlation and concordance. A main point is that voting focuses on the {*w, z*} plane in which "proportionality" stands for closeness to the diagonal rather than just having a ray through the origin, with $z = b\, w$ + e. Other texts on statistics may be vague on this (and may think about R). Some influences are the following:

- The *apportionment of seats* (italics) based upon votes involves some political philosophies that have been adopted by the national parliaments. Researchers on voting may have a tendency to remain with these philosophies when they *measure disproportionalities* (italics) from such apportionments too. Yet these activities (indicated by italics) have different purposes.
- There need not be a real distance between the voting literature and statistics, at roots, because (i) the Webster / Sainte-Laguë (WSL) apportionment philosophy compares with the Chi square, and (ii) the Gallagher index relates to the philosophy of minimising the sum of squared errors (Euclid), which philosophy is applied in the apportionment according to Hamilton / Largest Remainder. Perhaps this early historical linkage also caused the presumption that voting theory already "had enough" of what was available or relevant in the theory of statistics.
- The wasted vote concern votes for parties who receive no seats. The wasted vote clearly contribute to some disproportionality. The Kullback-Leibler (entropy) measure that compares vectors w and z still works when the w are the weights, but when we consider the condition of symmetry then it collapses on the zero in z.
- When equal votes shares {50, 50} translate in equal seat shares {50, 50} then this would be a proportional allocation, yet the conventional Pearson correlation coefficient collapses because of lack of dispersion in the vectors. (Thus **Section** 4 uses the cosine.)
- When vote shares {49, 51} translate into seat shares {51, 49} then this is major shift in majority. (Thus **Section** 4 uses the sign of the covariance.)
- There is also a zoo in statistics for measures for all kinds of different purposes on association, correlation and concordance. For voting researchers it might be difficult to find a match on the same kind of purpose.

The voting literature traditionally mentions the PR Lorenz and PR Gini too, but apparently in a different format than in Colignatus (2017b) that intends an improvement, see the abstract in **Appendix A** below. It may be observed that Lorenz and Gini are adequate for comparing proportional representation (PR) and district representation (DR). The present discussion on other measures then is not really required.

There may be more to be said on this, however. When I started looking at this issue, I wondered how the zoo in voting theory linked up to the zoo in statistics. This present discussion indeed improves clarity. It might be seen as a step forward that there is now a better distinction between focusing on the deviation of the slope from the diagonal versus focusing on other (more direct) criteria on w and z themselves. The discussion highlights where voting differs from other purposes in statistics: namely the requirement that $T = 1$. Any {*v, s*} proportionality conforms with a diagonal in the {*w,* z} space, but it helps to be aware of this. The discussion helps to see that the choice of a disproportionality measure apparently is not self-evident. Another point is that the PR Lorenz and PR Gini in the definition of Colignatus (2017b) require the ordering of the parties on the z / w values, while it might perhaps be advantageous (less sensitive to errors on this, keeping lists with parties in the same order e.g. for comparisons over years) when it would be possible to avoid such an intermediate step.

The following repeats explanations that are abundantly available in the literature, yet we should achieve more clarity, while a key point is the emphasis on the wasted vote. I will refer to Wikipedia – a portal and no source – for ease of access to the concepts used.

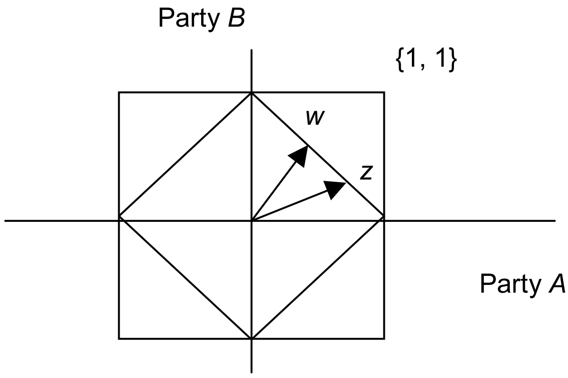
Before we look at measures of Kullback-Leibler and the Chi-square, it appears to be more fruitful to first discuss the unit simplex before looking at regression and the philosophy underlying the measure proposed in **Section** 4.

6. The unit simplex and its distance measures

6.1. The unit simplex

The normalisation $w = v / V$ and $z = s / S$ transforms to the unit simplex. Figure 5 gives the graphical interpretation for two parties, with approximately $w = \{40, 60\}$ and $z = \{70, 30\}$, for parties $\{A, B\}$. While a scatter diagram like Figure 3 has votes on the horizontal axis and seats on the vertical axis (which suggests the comparison of individual scores, like in the traditional disproportionality measures), the graph is now inside-out (and suggests the comparison of the vectors as wholes). The normalisation perhaps helps to see that the simplex might be the better way to look at the vectors. In this case, party *A* gets a higher share of seats than the share of votes. There would be proportionality in apportionment if the vectors would overlap (with anchor *w*). Votes and seats are located in the first Quadrant, and we might obliterate the other Quadrants, except that they may help to link up to distance measures developed more in general. If we would use Euclidean notions of addition and distance, then e.g. the vector addition of *w* and *z* would fall outside of the simplex, and thus these cannot be used.

Figure 5. Shares of votes and seats for two parties



6.2. Angular distance measure

A proper metric is given by the angular difference between the vectors. For example, the bisector of the angle between *w* and *z* generates a vector that can be normalised onto the unit simplex again, that can be said to be "halfway between".

For more dimensions, we get the $\theta = \mathrm{Cos}^{-1}[\mathrm{Cos}[w, z]]$, where the cosine is 1 when $\theta = 0$. We do not have to scale with *V, S* or *n* (the number of parties).

A comparison of $\theta_{Holland}$ and θ_{UK} would be meaningful. Their vectors will be in different planes, but the notion of an angle does not depend upon this. For a distance measure in [0, 1] we can use $d = \theta / (\pi/2)$ with θ in radians or $d = \theta / 90$ in degrees. This is a linear transform that neglects sensitivity.

(The situation becomes a bit awkward here. Colignatus (2009, 2015) proposes to get rid of π for the measurement of angles, and proposes to use the plane itself as a unit of measurement, with e.g. ½ turn for 180 degrees and ¼ turn for 90 degrees. When the angle θ between *s* and *v* is measured using the plane as a unit of account, then $d = \theta / (1/4) = 4\, \theta$ would be the transform for [0, 1]. See **Appendix D**.)

Angles and diagonal in the unit simplex should should not be confused with angles and diagonal in the scatter plot (each party scored separately), and these should not be confused with the angles and diagonal in the {*v, s*} space (vectors).

- The unit simplex helps to visualise $\theta = \text{Cos}^{-1}[\text{Cos}[v, s]] = \text{Cos}^{-1}[\text{Cos}[w, z]]$. The diagonal in the unit simplex is irrelevant.
- The diagonal in the scatter is relevant but then for slopes *T, b* and *p.* The cosine can be used there too as an "average slope" but perhaps with risk of confusion.
- The space of {*v, s*} creates (sub-) planes for particular observations. The diagonal 1 in that space has only a projection onto a plane, and need not be in the plane. The angles within the plane will be different from those with the diagonal in the whole space.
- Let $u = \lambda\, v + \mu\, s$ be on a plane in {*v, s*}, e.g. $\lambda = \mu = ½$. For more than 2 parties there need not exist weights such that $u = 1$. In this plane:

$$\theta v = \text{Cos}^{-1}[\text{Cos}[v, u]] = \theta w = \text{Cos}^{-1}[\text{Cos}[w, u]]$$

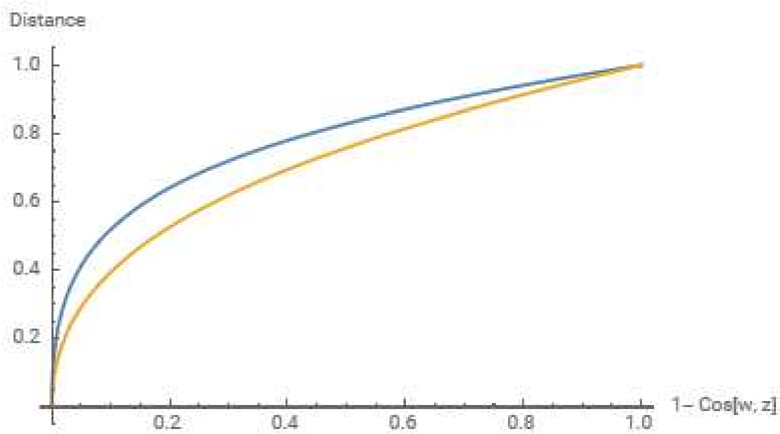
$$\theta s = \text{Cos}^{-1}[\text{Cos}[s, \text{u}]] = \theta z = \text{Cos}^{-1}[\text{Cos}[z, u]]$$

$$\theta = \text{Cos}^{-1}[\text{Cos}[v, s]] = \text{Cos}^{-1}[\text{Cos}[w, z]] = \pm\, \theta v \pm \theta s\ (\pm\, \pi) \text{ depending } \lambda, \mu$$

- With diagonal 1 of the {*v*, *s*} space: $\theta v1 = \text{Cos}^{-1}[\text{Cos}[v, 1]]$ and $\theta s1 = \text{Cos}^{-1}[\text{Cos}[s, 1]]$. Then only rarely $\theta = \pm\, \theta v1 \pm \theta s1\ (\pm\, \pi)$, for the relevant selection of signs.

The Weber-Fechner effect also applies to the angles. Small angles are important for voters but hardly noticed in a linear scale. A transform with logarithmic sensitivity is given in Figure 6. [5] One option is Log[1 + θ] and another is Log[1 + 4 θ], both normalised to 1 at $\theta = \pi/2$ (orthogonality). The latter option scales θ up to the full circle 2π. We already got the same effect with Sin in Figure 4. The display uses 1 – Cos, not as a metric but merely for switching the graph.

Figure 6. Distances on [0, 1] with logarithmic sensitivity to the angle between *s* and *v*



6.3. Evaluation on angle and slope measures

There is no fundamental difference between the angle measure in **Section** 6.2 and the proposed slope-diagonal deviation measure in **Section** 4 that transforms with Sin. Both are based upon *k* = cosine[*v, s*] and both are open to a Weber-Fechner transform. The only practical difference is whether we use $\text{Cos}^{-1}[k]$ to have an angle θ as an intermediate variable or whether we use Sin = Sqrt[1 – k^2]. Perhaps I am biased on this, but my impression is:

[5] Created in Wolfram Alpha with code Plot[{Log[1 + 4 ArcCos[1 - x]]/Log[1 + 2 Pi], Log[1 + ArcCos[1 - x]]/Log[1 + Pi/2]}, {x, 0, 1}, AxesLabel -> {"1 - Cos[w, z]", "Distance"}]

- The interpretation of the cosine *k* as a geometric average slope fits with the discussion on proportionality in terms of slopes.
- For *T* with other values than 1, standards on the slope of required proportionality *T* might rather not be formulated in terms of the angle.
- Likewise for voting we might think that the angle between *v* and *s* must be zero, but we rather tend to argue that the scatter should be along the diagonal, and we certainly do not tend to argue that the angles for *v* and *s* in the scatter must be 45 degrees for each. It might be an interesting innovation, though we can doubt whether it would improve the efficiency of the discussion.
- The discussion about the slope *b* provides a sound basis in regression (though this would not be needed since we already have the angle).
- The transform for the slope deviation from *b* to Sin = Sqrt[1 – *b p*] seems more tractable than the transform with arccos, log and its normalisation (though we might also use that transform again).
- The perspective on the cosine *k* as some kind of geometric average slope fits with the discussion on proportionality in terms of slopes, but might also be a somewhat confusing perspective. The comparison of cosine *k* with *T* might work for *T* = 1 but not for different values of *T,*which rather would be a tangens, and doesn't use all information that we have.

Thus one can see that the choice is one of preference only, and likely it is better to remain close to tradition. **Appendix B** has some residual comments on the unit simplex

7. Philosophy of the measure: error around the slope of 1

7.1. Symmetry and normalisation of the input: key step in the definition

Taagepera & Grofman (2003: 665-666) correctly require both symmetry and insensitivity to levels of *V* and *S.* Some first comments are:

- A regression of *s* given *v* would generate *S* / *V* when the slope *z* / *w* = 1. Thus a routine that has *s* and *v* as inputs should first normalise to *z* and *w.*
- Symmetry can be achieved by regressing both *z* given *w* (naturally) but also *w* given *z.*
- A regression *z* = *b w* can be regarded as a weighed regression *z / w* = *b* 1 with weights *w* or a weighed regression of 1 = *b w* / z with weights *z,* leaving out the 0 in *z.* Similarly for a regression *w* = *p z.* Thus there is ample scope for symmetry and there is a wealth of slopes to choose from. A potential development on the average that appears to be a digression is discussed in **Appendix E**.
- WSL with Sum[w (z / w – 1)^2] = Sum[(z – w)^2 / w] originated as the first expression, as a (natural) ratio measure (weighted by the votes). It is said that WSL favours the small parties because of the *w* in the denominator, but WSL actually attaches more weight to proportionality of the larger parties, perhaps at the cost of the smaller parties. Attaching value to the small parties would rather lead to Sum[(1 / w) (z / w – 1)^2].
- Symmetry would require that WSL with *z* / *w* is balanced by *w* / *z,* which however conflicts with our objective that the wasted vote are included in the measure. Having a party with 0 seats causes that *w* / *z* becomes undefined. It is conceivable to eliminate the 0 though.
- Lovell et at. (2015), Erb & Notredame (2016) and Quinn (2017) propose a statistic based upon the "log-ratio variance" or var[Log[*y* / *x*]]. They acknowledge that this is not symmetric and doesn't work for a 0 value. Obviously, Log[*y* / *x*] ≈ (*y* / *x* – 1) and the use of the variance introduces a square, so that the "log-ratio variance" is similar to the WSL. This links up to other approaches for the unit simplex, see **Appendix B**.

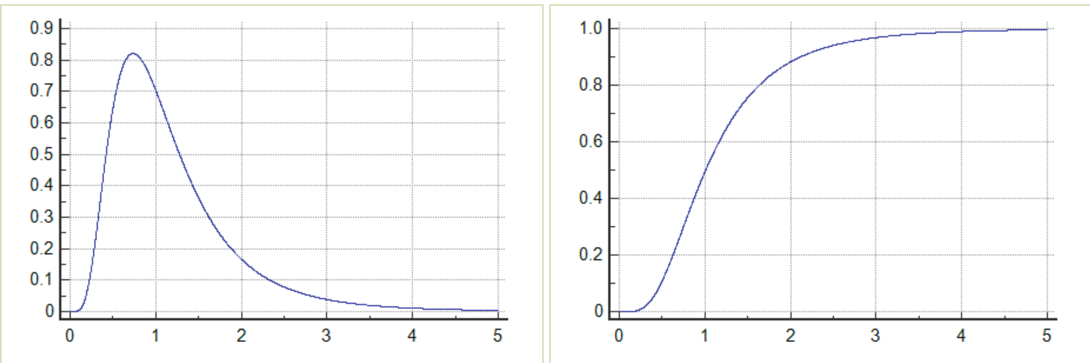
7.2. Variance ratio and the F-test

We may regard (unit) (diagonal) proportionality as a condition on the coefficients and then apply the F-test.

- The simple OLS estimate is $z = c + b\,w + e$, with sum of squared errors SSE[*c, b*] = $e'\,e$ = $n\ szz\ (1 - r^2)$ and $n - 2$ degrees of freedom, and *szz* the variance of *z.*
- The H0 hypothesis has $z = 0 + 1\,w + u$, with SSE[0, 1] = $u'\,u$ = Sum[$(z - w)^2$] and *n* degrees of freedom since no estimates are used here. Voting researchers will recognise a transform of Euclidean distance or the Gallagher index.
- The F-ratio test-statistic is *F* = *F*[*w,* z] = (SSE[0, 1] / *n*) / (SSE[*c, b*) / (*n* – 2)), for testing that the true situation is unit proportionality. We reject H0 when the ratio is too large according to some loss function. [6]
- The reasoning is that SSE[*c, b*] as OLS estimator already has minimal SSE, so that other values of the coefficients can only raise the ratio. If estimation renders *c* = 0 and *b* = 1, then the SSE are equal and *F* = (*n* – 2) / *n*.
- In Holland, the elections of 2017 for the House of Commons gave seats to 13 parties, so with the wasted vote lumped into one category, *n* = 14. With unit proportionality *F* = 12 / 14 = 0.86. Figure 7 [7] gives the density and cumulative distribution with degrees of freedom *n* =14 for the numerator (H0) and *m* =12 for the denominator (H1).

It is not very likely that one can say that the Dutch House of Commons started with the electoral result *w* and then allocated the seats *z* under the conditions that satisfy the assumptions of the *F*-ratio, which involves a degree of normal randomness. Thus the notion of testing is less relevant. But we can consider using *F* as a disproportionality index, since the present conceptual setting is sound, *F* works into the proper direction, and the probability is constrained to [0, 1]. A low value indicates closeness to unit proportionality, a high value indicates disproportionality.

Figure 7. F-ratio density and cumulative distribution for df {14, 12}



Some considerations are:

- Rather than *F* we would want to use the [0, 1] range, provided by the Fcdf.
- The part of the Fcdf below proportionality with *F* = (*n* – 2) / *n* is not used, and it would be impossible to reach 0 disproportionality. Thus we take a truncated distribution: *disp*[*x, n, m*] = Fcdf[*x, n, m*] / (1 – Fcdf[*m* / *n*, *n, m*]) for *x* = *F*[*w,* z] and *m* = *n* – 2.
- The Fcdf has the stark slope like the proposed SDD in Figure 4.
- The Taagepera & Grofman condition of symmetry however isn't met.
- The Taagepera & Grofman condition that measures of disproportionality should not be burdened with variables like the number of parties, gains in value. The condition is somewhat superfluous since we use vectors *v* and *s* that clearly have lengths. Yet the dealing with the relevant degrees of freedom per case is something that we rather avoid.

[6] (1) The following exposition is quite nice, though it uses a difference in the numerator: http://www3.wabash.edu/econometrics/EconometricsBook/chap17.htm. (2) Johnston (1972:28) has an explicit expression for entering hypotheses on *c* and *b,* but it is simpler to work directly with the SSE in above manner.

[7] https://www.medcalc.org/manual/f-distribution_functions.php

The above gives a clear conceptual setting but also motivates to look further.

7.3. (Dis-) Advantages of Pearson correlation

Disproportionality might also mean dispersion, as opposed to concentration, but proportionality rather links up with the notion of a linear relationship, which is precisely what the Pearson correlation coefficient focuses on.

- Thus 1 – R^2 (Pearson) would be an interesting point to start with. [8]
- The newer methods of DistanceCorrelation, [9] EnergyDistance [10] and the distance matrix [11] have been designed to discover dependence where Pearson fails, yet Pearson remains interesting precisely because of the notion of proportionality that leads to linearity.

The Pearson correlation however fails when some vector has no dispersion, which would be a common event in voting. A vote of {50000, 50000} and a seat allocation of {50, 50} would be perfectly proportional but Pearson correlation generates infinities. The reason for all of this is that Pearson looks at dispersion around the mean, which need not be relevant for (unit) (diagonal) disproportionality.

Concordance correlation (see the NCSS (undated) documentation) [12] has –1 for discordance, 0 for no concordance and 1 for full concordance. It generates 0 for what would be perfect proportionality as votes {50000, 50000} and seats {50, 50}, and generates infinities if we enter the data as percentages, for votes {50, 50} and seats {50, 50}.

Correlation however highlights that votes {49, 51} and seats {51, 49} are opposite. The Gallagher measure gives a value of 2, which one tends to regard as low, and which does not convey that there is such a dramatic reversal of majorities. We also get a Gallagher of 2 for votes {35, 34, 31} and seats {37, 34, 29} which looks a fair situation, given that it is a zero sum game. Only outcomes like {40,60} vs {60, 40} give a high Gallagher value of 20.

In other words, we have:

(i) the notion that proportionality T focuses on linearity through the origin
(ii) the notion that voting focuses on $T = 1$ in normalised space $\{w, z\}$
(iii) the fit of shares, in which {40, 60} and {60, 40} fit fairly well (Kullback-Leibler)
(iv) the implications for majority, in which the latter are quite disproportional.

It is an important insight: the sign of the correlation is a step towards checking whether an allocation reverses a majority. This forms a step towards handling power preservation more accurately.

While Correlation collapses for equal outcomes like {50, 50} since correlation requires dispersion in the variables, we can use the covariance or only its sign.

7.4. Regression through the origin: key step in the definition

With these potential sources of confusion out of the way, let us now look closer at what might be done with regression. Just to be sure: we are not *explaining* z by means of w, but we are merely *measuring* a degree of disproportionality, e.g. for cross-country comparisons. Statistically, it may be ill-advised to regress through the origin (RTO), since errors in modeling may show up in the constant which can be tested. For our purposes RTO is mandatory.

[8] https://en.wikipedia.org/wiki/Correlation_and_dependence
[9] https://en.wikipedia.org/wiki/Distance_correlation
[10] https://en.wikipedia.org/wiki/Energy_distance
[11] https://en.wikipedia.org/wiki/Distance_matrix
[12] Also https://en.wikipedia.org/wiki/Concordance_correlation_coefficient

If we have variables *x* and *y* then we say that these are proportional when $y = b\,x$, thus as a ray through the origin without a constant. Taagepera & Grofman (2003:663) mention that Cox & Shugart in 1991 already proposed to use *b* though for $y = c + b\,x$.

> "An advantage of [*b*] over other indices of disproportionality is that it indicates the directionality of the imbalance."

We want to get rid of the constant. Remarkably, portal Wikipedia (2017-08-07) (also) states: "Sometimes it is appropriate to force the regression line to pass through the origin, because *x* and *y* are assumed to be proportional." Thus *proportionality indeed*.

This gives us the OLS estimator of "regression through the origin" (RTO) $b = x'y / x'x$. See Eisenhauer (2003) for derivation and some comments, notably on calculating the R^2 and dangers of statistical packages.

A key point to be aware of is that with $y = b\,x + e$ then the sum of errors will not necessarily be zero. Thus *ymean* = *b xmean* + *emean,* so that on average there is an effect that compares to a constant, while there is no explicit constant in the equation.

Eisenhauer shows that a proper R^2 for RTO has $R^2 = b^2\, x'x / y'y = (y'y - \text{SSE}) / y'y$, while *y* are the levels and not the deviations from *ymean.* Substituting *b* we find: $R^2 = (x'y / x'x)^2\, x'x / y'y = (x'y)^2 / (x'x\ \ y'y) = k^2$, with *k* the cosine. Thus the proper *R* for RTO is the cosine. Subsequently the sine arises as a measure from $1 - k^2$. This thus differs from the Pearson R^2.

Thus we can take the RTO of *x* on *y* and find the other coefficient of determination. Essentially though, the SSE are calculated from the slopes, and basically we are formulating an error measure around these slopes too. This we do below.

Originally we worried that votes {50, 50} show no dispersion and thus would not allow the statistical estimate for *b.* However, the dispersion is only relevant for a regression with a constant, and not for a regression through the origin. (For a regression with a constant - perhaps for a sensitivity analysis – it remains useful to know that for {50, 50} the limit value of *b* is 0.)

7.5. Normalisation of the output: key step in the definition

Yet, *b* found by RTO is rather a proportionality measure and not a disproportionality measure.

- If $d = d[b]$ is a metric then $f\,d$ and $d / (1 + d)$ are metrics too, and the latter gives a normalisation to [0, 1].
- We can boost the sensitivity of lower values by a parameter *f,* so that $f\,d / (1 + f\,d) = 1 - 1 / (1 + f\,d)$
- A problem is that *d* loses the sign of *b,* which we just had discovered as a useful feature. This can be inserted separately. A zero covariance however might destroy relevant values of *d*. Thus we use the *sign* = If[sign[*b*] < 0, -1, 1].
- When we adopt the cosine then we can find similar adaptations.
- For Sin we can also use √Sin as a way to enhance sensitivity.

7.6. Restatement of diagonal regression, major axis, geometric mean functional relationship, or neutral regression

The following is required for an adequate overview. See Tofallis (2000) for an accessible overview, Lesnik & Tofallis (2005) for more, and Samuelson (1942) for skepticism w.r.t. the underlying model assumptions. [13] Following their references to Draper & Yang (1995)(1997), we there find the following. OLS regression of *y* given *x* with a constant gives coefficient *bols = sxy / sxx.* OLS of *x given y* gives *pols = sxy / syy.* Flipping the quadrant gives 1 / *pols* as an

[13] https://en.wikipedia.org/wiki/Total_least_squares#Scale_invariant_methods

alternative for *bols.* The diagonal regression (a.k.a. *major axis* or *geometric mean functional relationship* (GMFR) or *neutral regression*) estimate of the slope then is the geometric mean:

bdiag = sign[*sxy*] Sqrt[*bols* / *pols*] = sign[*sxy*] Sqrt[*syy* / *sxx*] = sign[*sxy*] *sy* / *sx*

cdiag = ymean – bdiag xmean

Draper & Yang (1995)(1997) state: "The GMFR estimate is obviously symmetric in that the interchange of *x* and *y* axes leaves the area unchanged." However:

- The numerical value of *bdiag* still depends upon which horizontal axis is taken.
- The intercept will generally not be zero, while we require zero for proportionality.
- Might that area be taken as a measure ? Consider *y* = *c* + *b x* + *e* so that the absolute vertical distance is Abs[*e*]. Then also *x* = (*y* – *c* – *e*) / *b* and the absolute horizontal distance is Abs[- *e* / *b*]. The product is Abs[*e' e* / *b*] = Abs[SSE / *b*]. The SSE is not the OLS SSE. We minimise over this expression (flipping axes to get rid of the Abs) and can confirm above outcome on *bdiag* and *cdiag.* The method is elegant and it is surprising that it is not used more often (notwithstanding Samuelson's skepticism). Anyhow, Draper & Yang already told us that Abs[*e' e* / *b*] apparently finds its (symmetric) minimum for *b* = *bdiag.* If I am correct then for nonnegative *bdiag*:

  Abs[*e' e* / *bdiag*] = 2 *n* (*sx sy* – *sxy*) = 2 *n sx sy* (1 – *r*)

  The question is whether this would be a disproportionality measure. It would be a transform of correlation. Observe that the LHS doesn't suffer from lack of dispersion in {50, 50} versus {40, 60} while the RHS tranform does. When the outcome would be 0 then we need not find this perfectly unit proportional (slope 1). As a measure it thus would suffer from the same drawback of the Pearson variation around a mean. This path dies in beauty.

(PM. Both OLS with a constant and the diagonal method differ from OLS through the origin (RTO). We intend to take *b = brto* and *p = prto.* Obviously we then might take *b* / *p* instead of above *bols / pols,* but then haven't resolved the issue of (numerical) symmetry.)

Tofallis (2000:9) has some useful comments on slope and correlation:

> "(...) reject RMA [diagonal, GMFR] on the grounds that the expression for the slope does not depend on the correlation between x and y. This is rather a strange objection because the correlation is a measure of the strength of the linear relationship and should be independent of the slope. The slope provides an estimate of the rate of change of y with x; why should this value be determined by the correlation (r)? After all, two regression lines can have the same slope but the data sets on which they are based can differ in the correlation; conversely, two sets of data can have the same correlation but have different regression slopes. It is conceivable that their objection may be grounded in the 'OLS conditioning' that most researchers are imbued with (since in OLS the slope is related to the correlation (r), according to: slope = r sy/sx). The eminent statistician John Tukey has indeed described least squares regression as a statistical idol and feels "it is high time that it be used as a springboard to more useful techniques" (Tukey, 1975)."

This should warn us about the interpretation of the OLS formula for the estimates of slope and *r.* The data allow these estimates (or summary statistics), but the data generating processes may have an entirely different structure.

7.7. Relation to the cosine distance: key step in the definition

Regressing through the origin, there is the slope *b* of z on *w,* thus *b* = *z' w* / *w' w,* and the slope *p* of *w* on *z,* thus *p* = *z' w* / z' z*.* Inverse values arise from flipping the quadrants.

The disproportionality measure *dmin* = 1 – Min[*b, p,* 1 / *b*, 1 / *p*] is symmetric and would be interesting since it uses all possible slopes generated by the data. The values of these slopes are nonnegative since *w* and *z* are in the first Quadrant and we force the line through the origin. It turns out that *dmin* is not so sensitive. We would like to see sensitivity. A difference of 1 seat in Holland is 1/150 = 0.67% and in the UK 1/650 = 0.15%, and would not be glossed over. Sensitivity can be enhanced by a transform, though. Somehow the "min" condition doesn't seem so appealing either, but it might perhaps only need getting used to. However, there is no clear linkup to the conditions of a metric, and thus this ends here.

There are other options. Slope *p* applies to the Quadrant with the axes flipped, so that we may compare *b* and 1 / *p* in the original Quadrant. A level distance is Abs[*b* - 1 / *p*], a relative expression is *b* / *(*1 / *p*) = *b p* and a geometric average Sqrt[*b* / *p*].

The relative expression is symmetric. It is the same as taking the two slopes by themselves in a more abstract manner. This gives *k* = Sqrt[*b p*] = *z' w /* Sqrt[*w' w* * *z' z*] or the Cosine *k*. [14]

- Thus the (relative, or "quadrant neglecting geometric average") slope estimator is also a measure of correlation. For us, however, the notion as a slope is relevant, and we define a deviation measure on it.
- More conventional is the CosineDistance 1 – *k*. However, the CosineDistance is not a metric, see Van Dongen & Enright (2012) and a counterexample. [15]
- Van Dongen & Enright (2012:4-5) find that the sine is actually a metric preserving distance, obviously defined as Sqrt[1 – *k*^2]. Also Sin[*x*]^*t* for 0 < *t* ≤ 1 gives a distance transform that we can use, though they do not explicitly say that it remains a metric.
- The cosine uses an average of the slopes, rather than selecting, like above Min condition, the worst outcome of either regression. Indeed, the CosineDistance generates low values where voters would rather see large values.
- Still, it seems reasonable to use the Cosine anyhow since it has a fair interpretation. Its lack of sensitivity can be adjusted by the transform, since this allows to increase the sensitivity in a more controlled manner. Geometry isn't sensitive to 1% difference but voters are.

Appendix C provides a follow-up on an angle here. PM. This weblog reviews some of the basic relations relevant for the present discussion. [16]

7.8. Relation to Hirschman-Herfindahl or Taagepera-Shugart ENP

Alternatively, the "quadrant-respecting geometric average" Sqrt[*b* / *p*] may be used. There is no real need for the Sqrt, though. We observe that *b* / *p = z'z* / *w'w* = *Nv / Ns*. In the same way as with diagonal regression the crossterm disappears. This gives a remarkable link to the Hirschman – Herfindahl [17] concentration indices *w' w* and z' z, and the Taagepera-Shugart 1989 measure [18] of the "relative reduction in the effective number of parties (ENP)" TS = 1 – *Ns* / *Nv.* Their use of the term "effectiveness" is not really defined, though. TS might generate negative values, which is no reduction but an increase, but that is reasonable when we see it as a a difference from the slope 1. TS is not symmetric. A potential disproportionality measure, with symmetric absolute error, thus based upon both *b* / *p* and *p* / *b*, is:

d = SymAbsTS = Abs[1 - *Nv / Ns*] Abs[1 - *Ns / Nv*] = $(Ns – Nv)^2$ / (*Ns Nv*).

It would still have to be normalised to [0, 1]. If all this would suffice then we would save on the calculation of the crossterms. I have not looked further into this. It is likely that there is useful information in the crossterms. SymAbsTS focuses on concentration while there might be (unit) (diagonal) disproportionality in the original dispersion.

---

[14] https://en.wikipedia.org/wiki/Cosine_similarity
[15] Adam Przedniczek gives the points {1, 0}, {1, 1} / Sqrt[2] and {0, 1}, see https://stats.stackexchange.com/questions/198080/proving-that-cosine-distance-function-defined-by-cosine-similarity-between-two-u/198103
[16] https://brenocon.com/blog/2012/03/cosine-similarity-pearson-correlation-and-ols-coefficients/
[17] https://en.wikipedia.org/wiki/Albert_O._Hirschman#Herfindahl-Hirschman_Index
[18] https://en.wikipedia.org/wiki/Effective_number_of_parties

Potentially, the regression through the origin (RTO) is also done in the diagonal regression fashion. We might say this of any regression. There is no reason to continue on this option. since we already found SDD.

7.9. Weighed regression

We encountered weighing in the WSL measure. In principle, we might measure each seat on its share of the vote, cf. single member districts. Summing over districts gives parties, and weighing by party shares thus makes sense. Parties also have the opportunity to collect errors to gain seats. The WSL measure isn't symmetric and meets with the problem of division by zero, but we can also use:

$$\text{symWSL} = \text{Sum}[w\,(z / w - 1)^2] + \text{Sum}[z^+\,(w^+ / z^+ - 1)^2] \text{ and } \{z^+, w^+\} \text{ has } z \text{ without } 0$$

We encountered weighing also in **Section** 7.1. The unweighed RTO $z / w = b$ 1 gives b = Average[z / w], so that this outcome can be seen as $z = b\,w$ with weights $1 / w$, that favour the small parties. This appears to be problematic, see **Appendix E**.

It would make some sense to have weighed regressions with weights w for $z = b$ w and $w = p\,z$. This would seem to tend to be more neutral to party mergers or splits.

For now, the SDD proposal has not opted for this. The cosine measure is straight from the first year linear algebra course and from econometrics textbook chapters without weights, and seemed simplest to work with at this stage. If SDD is adopted for disproportionality measurement, then one might keep in mind that parties react to rules of apportionment and not to measures of disproportionality in the research on electoral systems.

7.10. Thus the slope-diagonal deviation measure

These considerations thus generate the slope deviation measure, defined in **Section** 4.

The following sections first discuss other considerations before we look at numerical examples, since those considerations are relevant for judging the examples.

8. Apportionment, measuring, wasted vote

8.1. Definition of "wasted vote"

This **Section** starts with apportionment that minimises a minimand, then develops potential disproportionality measures based upon those minimands, and then supplements with additional information on statistics, parts of which we have already seen above.

With Turnout = Wasted Vote + Elected, or $V = W + Ve$, we have the vectors of perfect, though non-integer, allocations $a = S\,v / V$ and $ae = S\,ve / Ve$. The first includes the wasted votes and the latter excludes them. Standardly it is useful to collect all wasted votes into a single variable, not just to avoid a long list of unfamiliar names and zero seats (of far away countries). Issues on the number of (contending) parties should rather be treated separately.

Formulas will tend to be the same for both apportionment and the definition of a disproportionality index. It is tempting to choose for ease of exposition, which would mean the use of only *V, v* and *a* rather than *Ve, ve* and *ae.* Common texts on this subject choose for this ease of exposition and then appear to be confusing about the difference between apportionment and measures of disproportionality. Thus, we currently prefer some primness.

However, it would be a major principle for PR that the wasted vote should be represented by empty seats or be translated into a qualified majority. When we technically fill seats with blanco (dummy) representatives then we can still use the symbols V, v and a for both apportionment and a measure for disproportionality, see Table 5. After finding the relevant formulas, we then can substitute values of Ve, ve and ae, and then turn those blanco seats into 0, to fit current practice.

Table 5. Turnout = Wasted Vote + Elected: similarity of formulas

	$w = v / V$, $s[a = S\ w]$, $z = s / S$	$we = ve / Ve$, se[ae = $S\ we$], ze
Apportionment: s has no 0. Minimize a function on the error	(a) Wasted vote get seats with dummy candidates, not counted as 0. The allocation error is either $e = s - a$ or $e = s / a - 1$.	(b) Same formulas as the left but different base. The allocation error is either $ee = se - ae$ or $ee = se / ae - 1$.
Disproportionality measure = the minimand used in apportionment, or a separate standard	(c) Similar formulas as the above. Potentially one may subtract the value of the minimand found above, since the minimum above would be the (locally) optimal apportionment.	(d) When s is taken from (b), then the measure rises with the effect of excluding the wasted vote (and not using a qualified majority).

8.2. Apportionment (Jefferson, Webster, Hamilton)

Let us consider the situation in (a).

A main consideration is that seats s are taken in the range Floor[a] ≤ s ≤ Ceiling[a]. It is common to minimize a sum of squared errors SSE = Sum[e^2] for the given selection range. Yet, shall we take the error and SSE in levels or in ratios ? There is further no objective criterion for what would be unit proportional. Anything in the range Floor[a] ≤ s ≤ Ceiling[a] is a matter of choice, as only a is the proper proportion in a mathematical sense. Any choice might come with a paradox. The gains Γ = Max[0, $s - a$] for the electoral lucky compare to the losses Λ = Max[0, $a - s$] of those sent down to the Floor, with SSE = Γ ' Γ + Λ ' Λ. The value of dsa = Sum[Abs[$s - a$]] = Sum[Γ + Λ] gives a relative difference of allocation dsa / Sum[a], perhaps a minimal value that is unavoidable in apportionment, but not yet disproportionality.

HLR: This method is called the method of "Largest Remainder" or the method of Hamilton (HLR). With level error $e = s - a$, minimising SSE effectively means ordering the a – Floor[a] values (on size), and allocate the remainder seats to parties that would contribute most to the error. Lumping the wasted vote into "Other" may cause that it gets remainder seats.

(Let us reuse symbols f and c with a locally different meaning, and neglect a party index. Shifting from Floor f to Ceiling $c = f + 1$ means that the current SSE is reduced by $(a - f)^2$ and rises by $(c - a)^2$. The total change is $(c - a)^2 - (a - f)^2 = 2\ (c - a) - 1 = 1 - 2\ (a - f)$. Taking the highest values of $a - f$ causes the least increase of the SSE.)

WSL: This is the Webster / Sainte-Laguë (WSL) principle. The latter HLR disregards party sizes. For parties with 2 or 30 seats, an additional seat has different proportional meanings. Small parties are more likely to suffer relative underrepresentation. We could take $s / a - 1 = z / w - 1$ and weigh with the size of the votes w or attach more value to smaller parties with weights $1 - w$ or $1 / w$. These options are:

- minimand Sum[$w\ (z / w - 1)^2$] = Sum[$(z - w)^2 / w$]
- minimand Sum[$(1 - w)\ (z / w - 1)^2$] = Sum[$(z - w)^2\ (1 - w) / w$]
- minimand Sum[$(z / w - 1)^2 / w$] = Sum[$(z - w)^2 / w^3$] (exponent 3)

WSL uses the first approach. Its solution is to sort the parties on the Floor[a] / v ratio, and assign Ceiling values to the lowest outcomes, so that their position improves to Ceiling[a] / v. Unless the locus Floor[a] ≤ s ≤ Ceiling[a] is strictly imposed, other formulations of WSL might violate it. Lumping the wasted vote into "Other" may not avoid that this is treated as an overall small party, but the effect would be greater if they were not lumped together.

Paradoxes on HLR and WSL are: The choice of WSL over HLR is supposed to resolve the paradox that party sizes are overlooked. However, when one adopts WSL then it might be possible that a party with 10 seats might gain votes but lose a seat, while a party of 4 seats loses votes but gains this seat. This is only paradoxical when one would not see the underlying reasoning via the objective function.

JDH is Jefferson / D'Hondt (JDH). Gallagher (1991) finds that JDH apparently has the objective function to favour the largest party. Efficient tabular algorithms may look inversely at v / (k Floor[a] + 1) with k = 1 for JDH and k = 2 for WSL.

Given that WSL is a relative measure and we are considering PR anyway, it would seem that WSL would tend to be the better measure. However, for disproportionality measurement, it is not symmetric. However, a (political) property of apportionment is that it should not encourage parties to split up merely to gain seats. It is more commonly regarded as acceptable that parties are encouraged to join up and be rewarded by relatively more seats. On the other hand one would not want to overly encourage large parties at the cost of opposition. If the policy makers have these considerations then HLR would beat both WSL (potentially perhaps a bit better for small parties, but see Table 12) and JDH (large parties). That is, for apportionment.

The USA uses the Hill-Huntington (**HH**) method to allocate seats to State representatives in proportion to the population in the States, see Malkevitch (2002) [19] and (2017) and Caulfield (2010). There is little advice [20] to use this system also for the apportionment of seats to parties, and subsequently to turn it into a measure of disproportionality, since the current formulation assigns at least one seat to a state, while parties might belong to the wasted vote. Malkevitch (2017): [21]

> "Much of the theory here was developed by E.V. Huntington a mathematician who taught at Harvard University. Huntington had looked into the 32 ways that the inequality pi/pj > ai/aj (where the population of state i, is pi and the number of seats given state i is ai) could be rewritten by "cross multiplication." He worked out the different measures of "inequity" between pairs of states that could be used in this way. He observed that in a comparison between two states who had average district sizes of [large]100,000 and [small] 50,000 compared with [large] 75,000 and [small] 25,000, the absolute difference is the same [namely 50,000]. However, he thought that the inequity was "worse" in the second case because 50000/25000 is 2 while 50000/50000 is 1. The relative difference to Huntington seemed a better measure. (Relative difference between x and y being defined as |x -y|/min(x,y).) Of the absolute differences the two most "natural" are |pi/ai - pj/aj| which is optimal when Dean's method is used and |ai/pi -a/pj| which is optimal when Webster's method is used."

The topic of allocation of seats to states is different from the allocation of seats to parties, though, see also the seats for the Member States of the European Union. [22]

Observing the distinction between *apportionment* and *measuring disproportionality*, we may summarize current findings:

- *Apportionment* takes a decision following the minimisation of an objective function. Apportionment concentrates around the *locus of seat allocation*. This deals with small differences around a roughly proportional allocation, Floor[a] ≤ s ≤ Ceiling[a], in which it may happen that a party with votes may not get a seat due to missing the quota, but in which it should not happen that a party reaching the quota will not get a seat.
- *Disproportionality measures* contain the corresponding minimand, and are intended to judge about *wide differences*. The measures also must deal with cases, as this can

[19] http://www.ams.org/samplings/feature-column/fcarc-apportionii1
[20] https://en.wikipedia.org/wiki/Huntington%E2%80%93Hill_method
[21] Pairwise Equity in Apportionment (2017)
[22] https://en.wikipedia.org/wiki/European_Parliament#Members

happen in District Representation, that a party reaches the quota but still will not get a seat. DR has quite other considerations than PR.

- Apportionment focuses on parties that get a seat in the House of Commons, so that *s* has no zeros, while measures of disproportionality would rather include the wasted vote, consisting of invalid ballots and votes for parties that do not meet the quota, so that *s* has a zero (lumped together), since it obviously is disproportional to vote but not to be represented. The shares thus have different denominators, with the rule: Turnout = Wasted + Elected. (If a country would use a qualified majority such that the wasted vote can be replaced by *s* without 0, then this would be measured by rebasing to *s′* such that the wasted votes would have blanco seats, and allowing that this *s′* would not be integer.)
- Thus while apportionment would be perfect under its own rules, it would still show up as somewhat disproportional under its own associated disproportionality measure.
- For apportionment, a political principe would use HLR (on *Ve* and *ae*) for neutrality, to not encourage mergers and splitting of parties. (Though parties might use good polls.)
- For measurement one can assume that parties exist as they are. If we would restrict our attention to these traditional methods, then we could tend to use a relative measure as this better fits the relative character of proportionality itself. But a choice for WSL would not be seen as clearly better overall.
- This difference between apportionment and measurement doesn't need to create a great tension between what Parliaments do and what political scientists measure. (If so, one might subtract a mismatch on a baseline.)
- Measures are sensitive to listing all parties in the wasted vote separately or lumping them into a single "Other". It is advisable that lumping is the standard. For the Gini (in the definition in Colignatus (2017b)) lumping is immaterial but also preferable. The HLR / Gallagher index with its squares on levels would be higher with such lumping. Renwick (2015) is disappointed by the informational value of the Gallagher index w.r.t. the UK elections of 2015, but perhaps there is an issue on lumping.
- Above apportionment neglects issues on majorities. The notion of proportionality in apportionment is itself power preserving, but apparently in tradition it plays no role in the allocation of remainder seats. Disproportionality allows that a party / coalition with a majority in the votes would not get a majority in the seats, or that a party / coalition without such majority in the votes still gets it. Rules of apportionment that impose additional constraints on such preservation might be called "majority (unit) proportional". This would be a more sophisticated manner to improve the power preservation of systems that call themselves PR.

8.3. Disproportionality measures based upon apportionment

For measuring disproportionality, we would tend to use the *minimand* that is used in the apportionment rule, like Sum[$(s - a)^2$] or Sum[$(s - a)^2 / a$]. There is also the condition *Penalty* * (Max[0, Floor[a] – s]^2 + Max[0, s – Ceiling[a]]^2) for some penalty weight. We repeat some observations from the above but focus a bit more on this condition.

- (a) The Gallagher measure is a transform of the Euclidean distance, and corrects for the different number of seats in the various countries. With $a / S = v / V$ we get the minimand Sum[$(s / S - v / V)^2$]. The Gallagher measure then is Sqrt[1/2 Sum[$(z - w)^2$]] for observed values of *s* or *z* (and no longer the objective to apportion them). For a binary case with votes {w, $1 - w$} and seats {z, $1 - z$} with scalars now, the sum of squares is $(z - w)^2 + (1 - z) - (1 - w))^2 = 2 (z - w)^2$ so G reduces to Abs[$z - w$], the same value as the Loosemore-Hanby index (LHI)
- PM. The Gallagher measure treats values outside of Floor[a] ≤ s ≤ Ceiling[a] on an equal base. We may take allocation *a* as a fair estimate of both its floor and ceiling though.
- (b) The Webster / Sainte-Laguë method weighs the proportional errors by the votes, with Sum[$v / V * (s / v - S / V)^2$] = Sum[$(z - w)^2 / w$], so that Gallagher's terms are weighted by the inverses of the vote shares. The method might be more sensitive to values outside of Floor[a] ≤ s ≤ Ceiling[a]. It may be relevant here to stricter impose the penalty. If the penalty uses only squared values and no inverse weights then it would embed the measure within the wider range of the Gallagher measure.

- Gallagher (1991:47): "The Sainte-Laguë index (...) at the theoretical level is probably the soundest of all the measures (...). If it has a drawback, it is that it can be affected by the amount of information available on the fate of small groups." This refers to the lumping of all wasted votes into "Other". The latter can be advised as the standard.
- (c) Gallagher (1991:42) finds that the Jefferson / D'Hondt objective would be to maximise "the seats to votes ratio of the most over-represented party".

PM. Above measures all assume a single vote, with either the first preference or strategic voting. Systems like Single Transferable Vote (STV) allow the transfer of wasted votes to alternative parties. One would still record the first preference as the true vote, and the outcome after transfer might be measured as disproportional, though the idea of STV is that the outcome would be better than wasting the vote.

8.4. Statistical measures: Kullback-Leibler (KL) and Chi square

It remains important to be aware that *s* has been created by an apportionment on *ve* and that we now measure on *s* and *v.* The objective here is not to test apportionment but to see how such apportionment fares w.r.t. a measurement *standard*. In that case we may take greater liberty in choosing a disproportionality measure. Distance measures may be used in regression analysis or display, see Zand et al. (2015).

Kullback-Leibler (KL) [23] measures how shares or probabilities $z = s / S$ and $w = v / V$ compare. The divergence $D_{KL}[z \,||\, w]$ is also the relative entropy of *z* with respect to *w.* Taagepera & Grofman (2003:665) mention the measure simply as "entropy" and refer to an application by Pennisi 1998. It is not a true metric, since the outcome is asymmetric, but this would be okay for votes *w* and allocation of seats *z*. Belov & Armstrong (2011) have some results for normally distributed shares. Their lemma 2 has $D_{KL}[z \,||\, w] \approx ½ X^2$ for the *scaled* Chi square. [24]

The Pearson Chi square [25] is often used to test whether observed counts or frequencies are statistically different from the expected values. The allocation of the seats *s* is *observed* and we *expect* that it would follow the distribution of the votes *v* with expected value $a = S\, w$. H0 is that the vectors are the same, and we would reject H0 if the *p*-value is lower than a specified level. A requirement is that all $S\, w$ are at least 5 seats. With *n* the number of parties with seats, then the degrees of freedom would be $n – 1$, since Sum[$S\, w$] = S. It is debatable whether the assumptions for doing this test really apply though.

$$X^2 = \text{Sum}[\ (s – a)^2 / a\ ] = \text{Sum}[\ (s – S\, w)^2 / (S\, w)\ ] = S\ \text{Sum}[\ (z – w)^2 / w\ ] = S\ \text{WSL}$$

Apparently the Chi square is little used in the voting literature for this purpose, and this is a bit surprising, till we realise that it actually is identical to *S* WSL. This brings the comparison with the KL measure above to the fore. KL compares to WSL or a *scaled* Chi Square, while the standard Chi square test has *S* WSL. What to think about this issue on scaling ? Unfortunately this issue is not mentioned by Taagepera & Grofman (2003). However:

- Given that the Chi Square is a standard for comparisons, and given that we want to compare countries with different *S,* we apparently must also multiply the WSL with the total number of seats *S.* Gallagher (1991:46) table 4 scores countries on WSL but doesn't mention *S.* Including *S* would cause that the correlations with the other indices change. Holland (150 seats) and the UK (650 seats) have scores 2.1 and 23.5. To transform these into comparable standard X^2 we would get 150 * 2.1 versus 650 * 23.5. With such values, we would readily reject H0 that Holland or the UK have a unit proportional distribution.
- Thus the standard Chi Square with *S* WSL is so sensitive that it will generally cause us to reject H0 even for countries with PR like Holland. This makes the Chi square less useful.

[23] https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence
[24] https://stats.stackexchange.com/questions/9629/can-chi-square-be-used-to-compare-proportions
[25] https://en.wikipedia.org/wiki/Chi-squared_test

- Goldenberg & Fisher (2017) argue that WSL was based upon other considerations than the Chi square. This can readily be accepted. Also: WSL compares to KL.
- Thus WSL can still be used to compare countries, based upon WSL itself. WSL links up with the KL measure that also behaves like the scaled Chi square. (Belov & Armstrong (2011) lemma 2 was important for this insight.)

The assumptions of Pearson and WSL do not conflict in all respects. Pearson's assumptions only generate this particular format because the underlying (uniform) normality assumptions allow the derivation of an exact distribution, subsequently baptised with the name of X^2. One can start with the assumptions of WSL and then choose how to deal with the statistical issue. Either one adopts the statistical assumptions for hypothesis testing, or one merely uses the mathematical transform to create the disproportionality index in probability format. The probability transformation associated with the calculated X^2 still seems to be a tempting measure of (dis-) proportionality. Thus, this would not necessarily be taken as a statistical model but mainly as a useful transformation towards the [0, 1] range. A rising X^2 means a greater disproportionality. Thus $1 - p$[WSL] with p still taken as the $X^2[n - 1]$ would be the measure of disproportionality. Taagepera & Grofman (2003:664) mention that Mudambi 1997 proposed this. It remains to be seen whether it really can be used to compare countries (and their methods) over space and time.

(Remarkably, Pukelsheim (2014:130) puts the seats in the denominator, though correctly calls it a modification. He derives an objective function for a Member of Parliament based upon the number of voters that the MP represents, as Max v / s. Here, we rather look at an overall objective and use the format Min s / v.)

8.5. Correlation and cosine

Perhaps correlation is a more "objective" or "neutral" notion of association. It may be applied to v and s without the level corrections on V and S. The Pearson Correlation has the *distances from the mean*. We may however also consider measurement of the variables just from the origin (as they originally are).

Koppel & Diskin (2009) point to the Cosine as a measure that satisfies key properties. The cosine is also called "similarity". With vote vector v and seat vector s, Cos[v – Mean[v], s – Mean[s]] is the Pearson coefficient of correlation that reshifts to around the mean of the variables. This connection between cosine and correlation is a well established notion in the statistical literature, but potentially not in the voting literature.

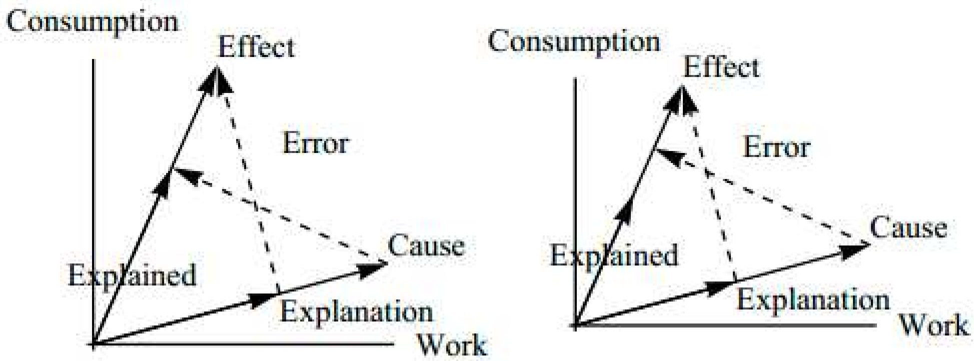
A potential measure is CosineDistance = 1 - cos[θ], [26] for θ = angle[s, v] that does not center the variables around their means. For the link of correlation and cosine to the use of the notion of (unit) "disproportionality" in the voting literature, I benefitted from Koppel & Diskin (2009), who propose to use this CosineDistance. The Wolfram Language introduced the CosineDistance = 1 - Cos in 2007. [27] See also the discussion by Zand et al. (2015). The cosine distance however is not a metric, see Van Dongen & Enright (2012).

Hill (1997) refers to Woodall 1986 in the *Mathematical Intelliger* who mentions that Dr J E G Farina proposed the cosine or correlation to measure the agreement of votes v and seats s. Hill has the remarkable suggestion to include {-v, -s} in the calculation of correlation to test on linearity. Ruedin, who refers to Koppel & Diskin (2009), has an online R-routine that he calls Farina. Likely it gives Cos and not 1 – Cos.

Figure 8 has been taken from Colignatus (2011:143) and shows how correlation and cosine relate.

[26] https://en.wikipedia.org/wiki/Cosine_similarity
[27] http://reference.wolfram.com/language/ref/CosineDistance.html

Figure 8. Correlation is no causation



The model is that the variation in Effect (Consumption) is explained partly by the variation in Cause (Work) and partly by unknown factors (error). The split is achieved by the projection of effect onto the cause ("Explanation"), while the dashed arrow from there to the Effect is the error. The *Effect* thus comes about by the vector addition of *Explanation* and *Error*. The perpendicularity makes these influences independent. Relative measures compare Explanation to the Effect, or Error to the Effect. A distance measure of 1 means zero Error, and that Cause and Effect fully overlap.

- For correlation the vectors are taken in deviation from their means (centered).
- A step further is to standardise them onto the unit circle, thus divide by the standard deviation, so that the centered vectors have unit Euclidean length.
- For the cosine the vectors are not centered and just taken as they are.
- A step further is to normalise them onto the unit simplex, thus divide by their length.

With two parties with vote shares {40, 60} and seat shares {60, 40), the correlation is –1 and the Cosine is 0.9231, giving a CosineDistance or 1 - Cos of 7.7%, which is rather insensitive. The Gallagher score is 20%. This actually also shows that measures of disproportionality might be insensitive to cases when a false majority is created. Comparing with the cosine for {10, 90} and {90, 10} the correlation remains –1 while the CosineDistance rises to 78%. The change from 7.7% to 78% seems proper, but an original problem was that 7.7% did not capture the false majority anyway. A idea is to use a measure so that the creation of a false majority is better indicated. A heuristic is to try to combine:

Correlation[*v, s*] * (1 – Cos[*v, s*])

A correlation index on votes {50, 50} collapses since correlation cannot deal with variables that have no variation. We can use the covariance then. However, a covariance of zero would be disinformative by itself, and destroys the information coming from the cosine. We are only interested in the switch to –1, and thus adapt the sign accordingly:

If[Covariance[*v, s*] < 0, -1, 1] * (1 – Cos[*v, s*])

Table 6 reproduces the data from Koppel & Diskin (2009), and includes the latter in the last column. The Cos measure is relatively unsensitive and we might apply the sign also to another index, like the Gallagher. PM. It seems curious to include cases with zero votes.

Table 6. Data from Koppel & Diskin (2009) including new last column

	Votes	Seats	D	R	G	G'	1-Cos	Sgn*(1-Cos)
1	{50,50}	{60,40}	.10	.10	.10	.14	.02	.02
2	{50,50,0}	{60,40,0}	.10	.07	.10	.14	.02	.02
3	{50,50,0}	{0,0,100}	1.00	.67	.87	1.00	1.00	-1.00
4	{25,25,25,25}	{30,30,20,20}	.10	.05	.07	.14	.02	.02
5	{25,25,25,25}	{35,25,20,20}	.10	.05	.09	.17	.03	.03
6	{50,50,0}	{50,0,50}	.50	.33	.50	.71	.50	-.50
7	{25,25,25,25,0}	{25,25,25,0,25}	.25	.10	.25	.50	.25	-.25

K&D observe that the Euclidean distance or Gallagher measure would be less sensitive to orthogonal outcomes, whence it would not measure disproportionality. Taagepera & Grofman (2003:666) have an example {25, 25, 25, 25, 0, 0} vs {0, 0, 0, 0, 60, 40}. In their story some Prince appoints people who did not get votes. Or, new elections fully replace parties. They regard this orthogonality as maximally "disproportional" too (or "largest change"). T&G still are inclined to stick with the Gallagher index, which generates 0.62 in their case while they want to see 1. But we are now also more acutely aware that the Prince also switches the majorities. The cosine distance for their Prince becomes 1 but we might wish to see –1.

Looking at the values of the cosine distance for the cases that Koppel & Diskin (2009) consider, we find that the sensitivity is not large. The cosine is not as sensitive to distinctions between *s* and *v* as voters would be. See also **Appendix C**.

8.6. Euclid (Gallagher index) and cosine

The Euclidean distance is a metric, and thus Euclid / √2 or the Gallagher index too. The index could be transformed to [0, 1] and made more sensitive, similar like the approach in **Appendix C**. Thus 1 – 1 / (1 + *f* G)) would match 0 with 0, and have a limit in 1 when G goes to infinity, while factor *f* would boost the low values for greater sensitivity. Perhaps such a transform should have been applied for true comparison of SDD with G in **Section** 9. However, there is no need for this, since SDD is better on theoretical grounds, and the voting literature is so used to G that one wishes to see it directly.

The relation between the Euclidean distance and the cosine gives (see also the discussion on the unit simplex in **Appendix B**):

||*z* – *w*||^2 = (*z* – *w*)' (*z* – *w*) = ||*z*||^2 + ||*w*||^2 – 2 *z' w.*
||*z* – *w*||^2 = 2 G^2 for Gallagher
||*z*||^2 = 1 / *Ns* or concentration, with *Ns* the effective number of parties ENP for *s*
||*w*||^2 = 1 / *Nv* or concentration, *with Nv* the effective number of parties ENP for *v*
cos[θ] = *z' w* / (||*z*|| ||*w*||) = *z' w* Sqrt[*Ns Nv*]
z' w = cos[θ] / Sqrt[*Ns Nv*]
Thus:
2 G^2 = 1 / *Ns* + 1 / *Nv* – 2 cos[θ] / Sqrt[*Ns Nv*]
2 G^2 = 2 (1 – cos[θ] / Sqrt[*Ns Nv*]) + (1 / *Ns* + 1 / *Nv* – 2)

G = Sqrt[(1 – cos[θ] / Sqrt[*Ns Nv*]) + ((1 / *Ns* + 1 / *Nv*) / 2 – 1)]

These corrections cause the G is a metric while 1 – Cos is not. Euclid / √2 reduces to Sqrt[1 - cos[θ]] when it is calculated on *v* / ||*v*|| and *s /* ||*s*||. Thus we can see that the Gallagher measure works in the same direction as 1 - cos[θ], though apparently not to full satisfaction as holds for the sine. As said, for the bipartite case, G reduces to LHI, and then G has the insensivity of the Dalton transfer, see **Section** 9.6.

8.7. A measure that uses the determinant

Let me mention the possibility discussed in Colignatus (2007) as well. The measure there is basically an ordinal measure, but the point is made that we tend to be forced to present nominal data in some order too. I have further not looked into a possible application to voting.

8.8. Can a uniform best measure be based upon tradition ?

Applying disproportionality measures involves counterfactuals. When country *C* uses objective function *O*, then a researcher using method *R* would have to explain why. A major consideration is to compare countries. If countries use the same apportionment then one might use that method. Comparing countries of PR and DR might involve the idea that there might be a more objective notion of disproportionality, which is somewhat dubious, even though this present paper suggests such a measure.

When we restrict our attention to the traditional methods of JDH, WSL and HLR, then we find:

- WSL, with its relation to the Chi square (divided by *S*), seems to generate the most interesting measure, namely the associated probability transform. This would not be regarded as if the statistical assumptions would apply, but it would merely be a transform that one can attach a [0, 1] meaning to. This requires the mention of the df.
- WSL has a form that suggests higher sensitivity to small parties, yet this form might be deceptive, and the discussion of Table 12 suggests that this causes the discussion: "when is small also small enough ?"
- If the menu card only contains the traditional measures, then there is no clear advice for a single measure. Seen from the world of the traditional measures, (unit) disproportionality has Byzantine aspects. Alongside the Gallagher of convention and the PR Lorenz and PR Gini for display and ease of communication, the discussion ***in this Section*** suggests that if voting research have a preference for WSL then they might consider (and then it still isn't in the [0, 1] or [-1, 1] range):

  Sign Sum[$(z - w)^2$ / w + *Penalty* / a * (Max[0, Floor[a] – s]^2 + Max[0, s – Ceiling[a]]^2)]

  in which Sign = If[Covariance[s, v] < 0, -1, 1].

9. Comparison of scores on theoretical cases

9.1. Comparison of bipartite cases around 50%

Table 7 gives scores of the disproportionality measures on theoretical bipartite cases. The second row states the range. The first three columns give label, votes and seats. The fourth column gives the proposed measure SDD with f = 2, or 10 √ Sin. Negative values indicate a switch in majority. Since there is risk of confusion on the scale, the fifth column multiplies by 10. The sixth column uses f = 1, or 100 Sin. The Gallagher measure is denoted by Euclid / √2. The Loosemore-Hanby index (LHI) is halve of the sum of the absolute differences.

Table 7. Scores of measures on theoretical bipartite cases around 50%

			10 Sqrt Sin	10 * left	100 Sin	Gini	WSL	Euclid / 1.4	LHI
Label	Votes	Seats	0 - 10	0 - 100	0 - 100	0 - 100	0 - inf	0 - inf	0 - inf
A	40	50	4.4	44.3	19.6	10.0	4.2	10.0	10.0
	60	50							
B	50	40	4.4	44.3	19.6	10.0	4.0	10.0	10.0
	50	60							
C	50	49	1.4	14.1	2.0	1.0	0.0	1.0	1.0
	50	51							
D	50	50	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	50	50							
E	50	60	4.4	44.3	19.6	10.0	4.0	10.0	10.0
	50	40							
F	51	49	-2.0	-20.0	-4.0	2.0	0.2	2.0	2.0
	49	51							
G	60	40	-6.2	-62.0	-38.5	20.0	16.7	20.0	20.0
	40	60							
H	49,000	49	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	51,000	51							
I	49,900	49	1.3	13.4	1.8	0.9	0.0	0.9	0.9
	50,100	51							

Comments are:

(1) Cases D and H are proportional and are generally recognised as such. Case H is only a test on normalisation to the unit simplex, but helps to compare to I.
(2) Cases B and F have already been discussed for Table 2, and one now has also the outcomes of the other measures.
(3) WSL, LHI and Euclid are somewhat difficult to judge upon given the infinite range. One rather compares values over the columns than over the rows. Perhaps we might hold that a Gallagher of 1 already is worrysome and a value of 10 alarming ? Potentially there might be little reason to compare electoral outcomes when we already know ahead that the results will be alarming.
(4) Cases A and B or E are symmetric, and this is generally respected except by WSL.
(5) Case C gives a small difference, and this gets more attention from SDD than the other measures. With a {50, 50} vote and a parliament of 100 seats, scoring an outcome {49, 51} with a low value 1, like the Euclidean distance does, might send the message that there is no reason to worry, except that it could also be much worse with {1, 99} of course. The geometric outcome of 2 by 100 Sin is low too, and one would prefer its square root.
(6) WSL is insensitive to the difference between C and D, and H and I. In I, WSL ≈ Euclid^2 / 50 so that the Gallagher outcome below 1 is squared and divided by a larger number, whence the result disappears in the lower digits, not shown. Multiplying WSL by 100 is not required since we already have done so by using percentages. WSL simply is insensitive to small differences.

9.2. Comparison of bipartite cases with a small party

Table 8 gives - which might be surprising given the title of this subsection and the header of the table - scores of measures on theoretical bipartite cases with a small party.

- The votes for cases J to N are {10, 90} and we look at a range of seats around 10.
- Case M is fully proportional. For reference disproportional case Q has also seats {10, 90},

Table 8. Scores of measures on theoretical bipartite cases with a small party

Label	Votes	Seats	10 Sqrt Sin 0 - 10	10 * left 0 - 100	100 Sin 0 - 100	Gini 0 - 100	WSL 0 - inf	Euclid / 1.4 0 - inf	LHI 0 - inf
J	10	0	3.3	33.2	11.0	10.0	11.1	10.0	10.0
	90	100							
K	10	1	3.2	31.7	10.0	9.0	9.0	9.0	9.0
	90	99							
L	10	5	2.4	24.1	5.8	5.0	2.8	5.0	5.0
	90	95							
M	10	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	90	90							
N	10	15	2.5	25.3	6.4	5.0	2.8	5.0	5.0
	90	85							
Q	30	10	5.4	53.9	29.0	20.0	19.0	20.0	20.0
	70	90							

Comments are:

(1) All measures pick up the move from disproportionality with 0 seats to proportionality with 10 seats, and then to disproportionality with 15 seats again.
(2) SDD is indeed more sensitive to disproportionality than Sin or Gini. When a small party with 10% of the vote doesn't get a seat, it means a score of 3.3 on a scale of 10.
(3) Perhaps 10 seats still is relatively much. There is nothing remarkable about this table, so that one would base decisions upon other results.

9.3. Comparison of bipartite cases with a tiny party

Table 9 looks at the position of a tiny party that might miss the threshold or that might get a bonus from some defectors of a large party.

Table 9. Scores of measures on theoretical bipartite cases with a tiny party

Label	Votes	Seats	10 Sqrt Sin 0 - 10	10 * left 0 - 100	100 Sin 0 - 100	Gini 0 - 100	WSL 0 - inf	Euclid / 1.4 0 - inf	LHI 0 - inf
R	1	0	1.0	10.1	1.0	1.0	1.0	1.0	1.0
	99	100							
S	1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	99	99							
T	1	2	1.0	10.2	1.0	1.0	1.0	1.0	1.0
	99	98							
U	5	0	0.7	7.1	0.5	0.5	0.5	0.5	0.5
	995	100							
V	5	1	0.7	7.1	0.5	0.5	0.5	0.5	0.5
	995	99							
W	5	2	1.2	12.4	1.5	1.5	4.5	1.5	1.5
	995	98							

Comments are:

(1) The pattern is remarkably similar to the case for the small party, though with some points.
(2) WSL has a reputation of being more sensitive to small parties, but doesn't really show this, except in case W, where it finds it quite disproportional that 0.5% of the vote generates 2% of the seats. This is also picked up by SDD, with 1.2 on a scale of 10.
(3) Case R: SDD is very sensitive to the case that 1% of the vote still is wasted. This gets 1 point on a scale of 10, or 10.1 on a scale of 100. The other scores also give 1, but some

also for the infinite range that is hard to judge. Compare with case J in Table 8 when 10% of the vote is wasted, and finds a SDD score of 3.3 on a scale of 10.
(4) There is the symmetry that wasting the vote in case R is judged as (almost) the same as in case T of having 1% of the vote and getting 2% of the seats. For now we can accept this symmetry, for it likely would be a different subject of how to evaluate vote waste.
(5) Similarly in cases U and V, 0.5% of the vote might either be wasted (0 seats) or be rewarded with 1 seat of 100 seats, and both cases are similarly disproportional.

9.4. Comparison of tripartite cases

Table 10 gives scores for theoretical cases for three parties.

Table 10. Scores of measures on theoretical cases: 3 parties

			10 Sqrt Sin	10 * left	100 Sin	Gini	WSL	Euclid / 1.4	LHI
Label	Votes	Seats	0 - 10	0 - 100	0 - 100	0 - 100	0 - inf	0 - inf	0 - inf
AA	48	2	5.3	53.2	28.3	20.6	17.1	14.7	17.0
	43	3							
	9	0							
AB	48	3	4.4	44.4	19.7	15.6	12.2	10.8	12.0
	43	2							
	9	0							
AC	50	0	-10.0	-100.0	-100.0	99.0	9900.0	85.7	99.0
	49	0							
	1	100							
AD	50	50	-9.3	-92.6	-85.7	73.5	2450.0	49.0	49.0
	49	0							
	1	50							
AE	50	60	4.3	43.2	18.7	10.4	4.7	9.5	10.0
	49	40							
	1	0							
AF	200	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	400	4							
	600	6							
AG	200	3	4.3	43.0	18.5	11.1	5.6	8.3	8.3
	400	4							
	600	5							

Comments are:

(1) Case AC shows the usefulness of the non-infinite range. Disproportionality cannot get very much worse, but the open ranges of WSL, Euclid and LHI still might suggest this. (We do not determine their ceilings in this manner, since there may be more parties.)
(2) Case AD was considered by Taagepera & Grofman (2003:672) in search of a notion of "halfway deviation from proportionality". Their definition of the Gini differs from the one given in Colignatus (2017b). They also allow zero votes, and this is replaced here by 1.
(i) A SDD value of 5.4 on a scale of 10 was found in case Q in Table 8, and may be close to the answer to their question what halfway would be.
(ii) T&G curiously write values of Euclid and LHI with a %-sign but these are not percentages. Indeed, these measures generate values close to 50 but this cannot be interpreted as 50%. The SDD score indicates strong disproportionality.
(iii) For comparison, case AE gives a more proportional outcome, and SDD provides more guidance here than the other measures.
(3) Cases AA and AB have been taken from Hill (1997:8), who argues for Single Transferable Vote (STV) - rather than doubling the number of seats, based upon the observed disproportionality:

> Even within strictly party voting, the first-preference measures are unsatisfactory. Consider a 5-seater constituency and several candidates from each of Right, Left and Far-left parties. Suppose that all voters vote first for all the candidates of their favoured parties, but Left and Far-left then put the other of those on the ends of their lists. If the first preferences are 48% Right, 43% Left, 9% Far-left, all the measures will say that 3, 2, 0 is a more proportional result than 2, 3, 0. Yet STV will elect 2, 3, 0 and that is the genuinely best result, because there were more left-wing than right-wing voters. There is no escape by comparing with final preferences, after redistribution, instead of first preferences. That is merely to claim that STV has done well by comparing it with itself. Our opponents may sometimes be dim, but I doubt whether they are dim enough to fall for that one.

(4) Cases AF and AG are also from Hill (1997:7), whose issue with correlation now would be answered:

> For example with votes of 200, 400 and 600 and the proportional 2, 4 and 6 seats we get a correlation of 1.0, but the non-proportional 3, 4 and 5 seats equally get 1.0 as those points also fall on a straight line. To get a suitable measure we also need to include the same numbers over again, but negated. Thus 200, 400, 600, −200, −400, −600 with 2, 4, 6, −2, −4, −6 gives a correlation of 1.0 as before, but 200, 400, 600, −200, −400, −600 with 3, 4, 5, −3, −4, −5 gives only 0.983 demonstrating a less good fit.

9.5. Comparison of quadripartite cases

Table 11 gives some cases with four parties. Comments are:

(1) Cases AAA and AAB show a marginal change in disproportionality given an already high basic level. LHI is insensitive to the change. Overall it comes as more natural to attach meaning to SDD and Gini as opposed to the other scores.
(2) Case AAC shows the benefit of the sensitivity to majority reversal. Obviously, the sign can also be introduced for the other measures.

Table 11. Scores of measures on theoretical cases: 4 parties

Label	Votes	Seats	10 Sqrt Sin 0 - 10	10 * left 0 - 100	100 Sin 0 - 100	Gini 0 - 100	WSL 0 - inf	Euclid / 1.4 0 - inf	LHI 0 - inf
AAA	25	30	4.4	44.3	19.6	10.0	4.0	7.1	10.0
	25	30							
	25	20							
	25	20							
AAB	25	35	4.9	48.8	23.8	12.5	6.0	8.7	10.0
	25	25							
	25	20							
	25	20							
AAC	33	33	-8.6	-86.2	-74.2	54.8	1122.0	33.0	33.0
	33	33							
	33	0							
	1	34							

9.6. Comparison on the Dalton transfer

We can judge measures on their minimands and scores on situations, and situations on scores, but also look at the marginal change of taking one seat from a party and apportioning it to another. This is callled "Dalton's transfer". Taagepera & Grofman (2003:667) (T&G) give an example. When a seat is transferred from a richer party to a poorer one, the (unit) disproportionality should decrease. Conversely, an increase.

I am not comfortable with this principle, since apportionment would be restricted to the Floor-Ceiling locus, and it is not clear what we are actually doing when shifting a seat, as this might be in any direction from a local optimum in the locus. I would also prefer mathematics to get at a general statement. Indeed Goldenberg & Fisher (2017) (G&F) provide some relevant formulas on WSL and Dalton's principle. I have not tried to see what a symmetric version would do, given the problem of division by zero.

Table 12 contains the T&G example. Case DA and a transfer of 1 seat from 60 to 15 causes DB. Case DA and a transfer of 1 seat from 15 to 60 causes DC.

(i) SDD and Euclid show differences only in the third digit.
(ii) T&G have a different definition of the Gini but the outcomes here are the same.
(iii) The order of disproportionality for SDD and Euclid is DC > DA > DB, which fits Dalton's principle. Gini and WSL have DB > DA > DC, i.e. reversed.
(iv) The crucial distinction is that SDD is symmetric by design and the PR Gini is asymmetric by design. They have different objectives.
(v) T&G explain the Gini outcome by looking at the *z* / *w* "advantage ratio". Their conclusion: "In the numerous cases where the same party (often the largest) has both the largest difference and also the largest ratio, of course, [Euclid] and Gini pull in the same direction (...)." [though here they are reversed] and "Yet the mind baulks at adding seats to a party that is already 50 percent overpaid and calling it a reduction in disproportionality. Maybe this is the crucial difference between the notions of 'deviation from proportionality' (stressing the differences) and 'inequality' (stressing the ratio)."
(vi) I wonder whether the distinction between level and ratio really matters. It is rather that the Gini is asymmetric because of its very purpose. I can concur with this: the Gini is not a measure on (unit) disproportionality (slope-diagonal deviation) but one of inequality as defined by the Gini. A It remains true that the present implementation of the Gini sorts on the "advantage ratio", and that the cumulation of disadvantaged values causes that they affect the subsequent cumulative values.
(vii) Overall, this example by T&G is enlightening on this comparison of Gini and SDD.

G&F rightly point out that T&G claimed that WSL satisfies Dalton's principle (in levels) while it might sometimes do and sometimes not do.

(viii) G&F point to the UK general election 2010 as a counterexample, while they might also have refered to the very T&G example itself.
(ix) These issues should rather be judged within the Floor-Ceiling locus of apportionment, see **Section** 8.8. If it is not clear whether the base situation is proportional then it may not be clear what the effect of shifting a seat might be. (For proportionality along the ray through the origin, one might adapt the number of seats too.) Shifting 1 seat outside of that locus of apportionment with a given number of seats might be awkward, for one cannot claim that a party has a right for a seat outside of that range. With a proper metric, though, we may consider that a situation outside of the locus must be disproportional, and a more distant situation even more (unit) disproportional.
(x) There is also symmetry, and not just the distinction between levels and ratios.
(xi) That said, in the example of Table 12, case DB gives the smallest party the most seats. SDD judges this as the least disproportional given the situation on the other parties, while WSL judges it as most disproportional. It is often said that WSL has greater sensitivity for smaller parties, but case W in Table 9 shows also the intolerance of overrepresentation by small parties.

Table 12. Taagepera & Grofman (2003:667) on Dalton's principle of transfer

			10 Sqrt Sin	10 * left	100 Sin	Gini	WSL	Euclid / 1.4	LHI
Label	Votes	Seats	0 - 10	0 - 100	0 - 100	0 - 100	0 - inf	0 - inf	0 - inf
DA	50	60	5.3	52.9	28.0	16.5	10.1	13.2	15.0
	40	25							
	10	15							
DB	50	59	5.3	52.8	27.9	17.1	10.8	13.1	15.0
	40	25							
	10	16							
DC	50	61	5.3	53.1	28.2	15.9	9.6	13.5	15.0
	40	25							
	10	14							

10. Scoring on the Taagepera – Grofman criteria

Taagepera & Grofman (2003) (T&G) give some criteria for (unit) disproportionality measures, and score the more traditional measures on them. Let us check how SDD does. We already met some criteria like "being a metric" in **Section** 4.3. T&G do not explicitly say so, but they have similar conditions, and if the Dalton transfer can be translated into triangularity, then they require a metric. It is useful to run through the T&G list afresh.

(1) Yes. Informationally complete: makes use of the v and s data for all parties. (This excludes measures that take only the two largest parties.)
(2) Yes. Uniform: uses the data uniformly for all parties.
(3) Yes. Symmetry.
(4) Yes. Range. (Nonnegative.) Varies between 0 and 1 (or 100 percent). (i) The amendment is to prefer [1-10] though. (ii) The amendment is to allow for the negative sign on majority switches.
(5) Yes. Zero base. Has value 0 if $s = v$. (The comment though is that commonly s has a zero because of the wasted vote, and we avoid cases with a 0 in v. We have $s = v$ only in theoretical cases or in comparing s and Ve in Table 5.)
(6) Yes. Has value 1 or 100 percent for full disproportionality, when a positive vote matches with 0 (excluding the wasted vote) or conversely. (See their example of the Prince.)
(7) There is a fair effect on Dalton's transfer, see **Section** 9.6. This is not worked out fully though.

(8) Yes. Does not include the number of parties. This is a curious criterion since the vectors have a length that gives the number of parties. Yet, we can imagine that there are some practical considerations involved, that cause a desire to not be explicit about this, like the degrees of freedom in the Chi square distribution.
(9) Yes. Insensitive to lumping of "residuals" (individual parties in the wasted vote). This lumping is both a practical consideration and one of fundamental principle, e.g. on the wasted vote that is collected into "Other". (i) T&G claim that the Gini is sensitive to lumping, but they use a different definition, and the PR Gini in Golignatus (2017b) is insensitive to lumping. (ii) Taagepera & Grofman (2003:668) give the example reproduced in Table 13. Cases XA and XB differ because only XB has a wasted vote, so that SDD produces a higher score. Normally, Euclid is sensitive to lumping, but in this case the percentage difference in XA for small parties 5 is zero, and in XB the difference 1 - 0.1 = 0.9 becomes even smaller when squared. Thus the example is not as strong as can be made. But there is no need to construct another example since the principle is clear. As a standard, it remains advisable to lump the wasted vote and invalid votes into one category.
(10) Yes. Is simple to compute. The notions of cosine and sine are highschool issues, see **Appendix D**. For vectors there is a well established theory of at least some 100 years. The calculation can be done in a simple spreadsheet. Taagepera & Grofman (2003:670) argue that "[Euclid] is difficult to calculate", and give it a score of 0.5 on simplicity. It is hard to believe that they are really serious on this. Their yardstick is LHI of course.
(11) Yes. Is insensitive to a shift from fractional to percent shares. It is somewhat remarkable that this needs to be mentioned. (Some attention to scale adjustment is only fair.)
(12) Yes. Input depends only upon shares $w = v / V$ and $z = s / S$ without use of V and S. (However, researchers would tend to require full data, also for the relation to apportionment. Thus the criterion is practical only in a limited sense. For example, the UK with 650 seats has a greater potentiality to get a proportional fit than Holland with 150 seats. (Check the value for PR+ in Table 4.) Also, it is better to store the data in the form of integers v and s rather than in fractional form (with at least 6 digits precision too).)

Table 13. Taagepera & Grofman (2003:668) on lumping small parties in "Other"

Label	Votes	Seats	10 Sqrt Sin 0 - 10	10 * left 0 - 100	100 Sin 0 - 100	Gini 0 - 100	WSL 0 - inf	Euclid / 1.4 0 - inf	LHI 0 - inf
XA	400	50	4.8	47.9	22.9	14.0	7.5	10.0	10.0
	300	30							
	200	10							
	50	5							
	50	5							
XB	400	50	5.3	53.3	28.5	29.1	97.5	12.0	19.0
	300	30							
	200	10							
	90	0							
	1	1							
	1	1							
	1	1							
	1	1							
	1	1							
	1	1							
	1	1							
	1	1							
	1	1							
	1	1							

To summarise:

- Overall, SDD scores perfectly well on these criteria, except that the difference transfer hasn't been worked out mathematically yet, unless it suffices that SDD is a metric.
- Overall, the Taagepera & Grofman (2003) criteria are mostly technical rather than criteria on content. The difference transfer and lumping are the most important on content. The difference transfer was important to highlight the distinction between symmetric SDD and asymmetric WSL and PR Gini.
- Remarkably, the regression coefficient recieves an overall low score. When one wonders why this is so, then the chain of reasoning and the elimination of the deficiencies can lead to the creation of the SDD.
- We agree with their conclusion p673: "Favouring some other index for any one of these topics would need justification to counterbalance the shortcomings listed here."

11. Conclusion

The slope-diagonal deviation (SDD) measure for (unit) disproportionality defined in **Section** 4 is rather simple. Perhaps it might look a bit complex but then this is only because of the surrounding zoo of indices both in voting theory and statistics. It is remarkable that SDD hasn't been formulated early at the start when researchers started looking into electoral systems in a quantified approach. (One thinks about Sainte-Lague (1882-1950) and Hill (1860-1938) and Huntington (1874-1952), though the latter authors focused rather on seats per state than seats per party. [28])

As an econometrician who has been familiar with the relationship of cosine and correlation since the first year linear algebra course, I found the literature in voting theory about disproportionality measures bewildering. It comes across as if these researchers try to reinvent statistics without awareness of the wealth of statistical theory. It already starts with the confusing use of the word "proportionality" for only *unit* or *diagonal* proportionality. In that manner, voting researchers also create a barrier for statisticians who want to understand what their problem is. It should help to explain at the very start:

- Any proportional relation $s = T\ v$ reduces in $\{w, z\}$ space to diagonal $z = w$ because $S = 1's = T\ 1'v = S$. Division namely gives $z = s / S = T\ v / S = w$.
- The notion of unit proportionality as in the line $y = 1\ x + 0$ is a mathematical concept, but in statistics with dimensions we can always rebase the variables, so that there is no natural base for 1. Deviations from $z = w$ however make sense.
- The proper tool to use here is regression through the origin (RTO), with the coefficient of determination $R = Cos$. Regression of {40, 60} vs {60, 40} doesn't give $z = 1\ w + e$.

Voting research has a wealth of insights and its texts have reason and rationality. However, when a text presents a conclusion that such and such measure *M* behaves such and such, then this may be true, but it becomes irrelevant once it is (already) clear from statistics that this measure *M* can be discarded. **Section** 9 above has a Byzantine structure of comparing the proposed SDD with inadequate measures, and the Section is actually superfluous once one understands what SDD does. Yet for voting researchers the Section will be highly relevant because it relates to how they understand the world. I have tried to find a measure that is close to this world of thinking (like a Gallagher transform to [0, 1] with increased sensitivity) but decided that I should not sacrifice statistics for this otherwise noble purpose.

A key point for modesty remains that voting research comes from the tradition exemplified by Jefferson, Webster and Hamilton that votes have to be apportioned into seats, and that the choice is not self-evident. It is important that this tradition exists, and we can only respect that it apparently runs its own course with its own traditions, like statistics itself has its traditions. Kenneth Arrow's confusion on the interpretation of his Impossibility Theorem – see Colignatus (2014) on the distinction between voting counts and decisions – is another example how research might get lost on fundamental ideas. Or **Appendix D** is sobering on the education of mathematics and its research for the last 5000 years. Fortunately, voting in a free world may help prevent that tradition becomes dogma.

[28] http://www-history.mcs.st-andrews.ac.uk/Biographies/Huntington.html

12. Appendix A. Abstract of Colignatus (2017b)

QUOTE

The Lorenz curve and Gini coefficient are applied here to measure and graph disproportionality in outcomes for multiseat elections held in 2017. The discussion compares Proportional Representation (PR) in Holland (PR Gini 3.6%) with District Representation (DR) in France (41.6%), UK (15.6%) and Northern Ireland (NI) (36.7%). In France the first preferences of voters for political parties show from the first round in the two rounds run-off election. In UK and NI the first preferences of voters are masked because of strategic voting in the single round First Past the Post system. Thus the PR Gini values for UK and NI must be treated with caution. Some statements in the voting literature hold that the Lorenz and Gini statistics are complex to construct and calculate for voting. Instead, it appears that the application is actually straightforward. These statistics appear to enlighten the difference between PR and DR, and they highlight the disproportionality in the latter. Two conditions are advised to enhance the usefulness of the statistics and the comparability of results: (1) Order the political parties on the ratio (rather than the difference) of the share of seats to the share of votes, (2) Use turnout as the denominator for the shares, and thus include the invalid and wasted vote (no seats received) as a party of their own. The discussion also touches upon the consequences of disproportionality by DR. Quite likely Brexit derives from the UK system of DR and the discontent about (mis-) representation. Likely voting theorists from countries with DR have a bias towards DR and they are less familiar with the better democratic qualities of PR.

UNQUOTE

PM. References that are not in the article: Taagepera & Laakso (2006) and Laakso & Taagepera (2007) plot in the $\{v, s / v\}$ space, calling this a "proportionality profile". This is close to using the Gini.

13. Appendix B. Some comments on the unit simplex

13.1. Gallagher and concordance

See also **Section** 8.6 on the relation of Gallagher and the cosine. There we use a different notation. See Figure 5 for the unit simplex.

The Gallagher index has the base in Euclid Sum$[(z - w)^2] = (z - w)' (z - w) = z'z - 2z'w + w'w$. Given that this is nonnegative, we have $2z'w \leq z'z + w'w$.

For voting only Quadrant I with nonnegative vectors is relevant. There we can define "Origin concordance" and its implied distance as:

$$0 \leq \text{ OC} = z'w / (½\,(z'z + w'w)) \leq 1$$

$$0 \leq \text{ OCDistance} = 1 - \text{OC} \leq 1$$

- The OC and OCDistance are symmetric, with the proper range, and OCDistance = 0 when $z = w$. I have not checked whether OCDistance is a proper metric though.
- Cosine and OC take variables from the origin while Correlation and Lin's concordance (LC) (**Section** 7.3) [29] take them around their means.
- The Cosine has a geometric average in the denominator and OC an arithmethic.

[29] https://en.wikipedia.org/wiki/Concordance_correlation_coefficient

Looking at the other quadrants, we find that the lower zero bound is also met in Quadrant III, when the negative signs cancel each other. In Quadrants II and IV the minimal value –1 can arise, for example for w = {-1, 0} and z = {1, 0}. (This is not an example for voting.) For general {x, y} on the unit simplex we have:

$$-1 \leq \text{OC} = x'y \,/\, (½\,(x'x + y'y)) \leq 1$$

It is a bit tedious to figure out how to translate this into a distance. For nonnegative OC the transform 1 – OC already was fine. What to say about a value OC < 0 ? For w = {- ½, ½} in Quadrant II and z = {½, -½} in Quadrant IV, we get (-¼ + -¼) / (½ ((¼ + ¼) + (¼ + ¼))) = -1. Thus we have a directed distance:

$$-1 \leq \text{OCDistance} = \text{If[OC < 0, OC, 1 – OC]} \leq 1$$

I have not looked into this further (for the potential use of a (unit) disproportionality measure), because: (i) the similarity of arithmetic and geometric averages suggests that it would not matter so much, (ii) the cosine is already well established, and I stumbled upon this OC while writing this and have not looked whether there already is some literature on it (perhaps under some other name) (and I might have plainly forgotten about it, simply because it wasn't so relevant until now) (there is already a relation to cosine and the harmonic mean of the squares).

13.2. Aitchison geometry

Pawlowsky-Glahn et al. (2007) provide an accessible and enlightening overview for the statistical analysis of "compositional data" and the Aitchison geometry. The "compositional data" are the vectors in the unit simplex, a.k.a. relative information, parts per unit, proportions, percentages, and so on. The issue may be more complex than for electoral systems, and for example contains the problem when researcher *A* measures a sample compound on four variables and researcher *B* does this one month later for the same sample on three variables: so that it can appear that a positive correlation in the first study turns into a negative correlation in the second study.

The Aitchison geometry subsequently uses positive vectors and logarithmic transforms (with values [0, infinity], also called the log-ratio approach. This is not useful for voting in which the wasted vote has 0 seats. The log-ratio compares to WSL that we found has drawbacks.

(PM. The short overview text doesn't mention the angle metric. I have not looked into this further.)

14. Appendix C. Using Abs[k – 1 / k] and another transform

The following continues the discussion of **Section** 7.7.

First observe that if k is the cosine for the nonnegative quadrant that we are looking at, then Tan = Sin / Cos = Sqrt[1 – k^2] / k = Sqrt[1/k – k] / Sqrt[k]. The denominator might not always be needed. The cosine remains interesting because it already is in the range [0, 1] while other measures are not, and then require a transform.

If we use $k = b / p$ then we may consider using d = Abs[k – 1 / k] as a symmetric expression of deviation around slope 1.

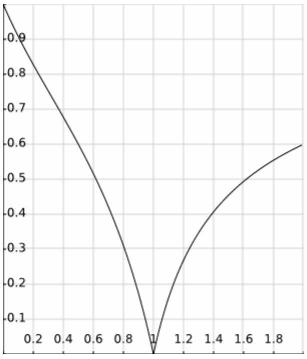
For k > 0 we have Abs[1 – k^2] = Abs[k – 1 / k] k and this will work in the same manner as Sqrt[1 – k^2]. Only for Sin it is established that it is a metric.

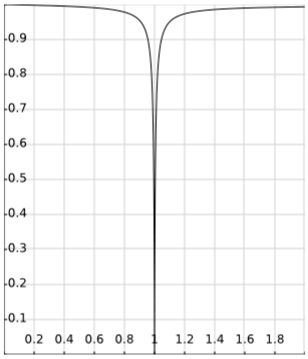
For a transform of this d we can use a stable function with a fixed parameter f as a norm for sensitivity, $f\,d$ / (1 + $f\,d$), which gives *disp*[k] = 1 – 1 / (1 + f * Abs[k – 1 / k]). Relevant values for f are 1, 10, 100 and 1000. Figure 9 shows *disp* for d = Abs[k – 1 / k] for values of $0 \leq k \leq 2$

and factors f = 1 and 100. [30] An increasing sensitivity will quicker decide upon higher disproportionality.

NB. This approach assumed $k = b / p$, while the body of the text converged on selecting k = *cosine*[*v, s*], which has only the range [0, 1]. At first it is somewhat remarkable that the graphs for this selection and for the selection of the cosine and sine are quite similar, but the logic behind both approaches is also quite similar.

Figure 9. *disp*[*k*] = 1 – 1 / (1 + *f* * Abs[*k* – 1 / *k*]), 0 ≤ *k* ≤ 2, for *f* = 1 or 100





15. Appendix D. Xur & Yur

Colignatus (2009, 2015) and (2011) show how sine and cosine can be re-engineered. Figure 10 contains two figures taken from there. The text uses $H = -1$, so that $x^H = 1 / x$.

- The unit of angular measurement is the whole plane itself.
- Thus angles have a [0, 1] domain on the unit circumference circle. This is also called the "unit measure around" or "unit meter around" (UMA). This compares to 360 degrees (artificial, number of days in a year) or 2π radians (unit radius circle).
- Sine and cosine can also be discussed in terms of xur and yur.
  - They derive from both a ratio and the unit radius circle, and their name can use "ur"
  - The cosine gives the horizontal value or *x,* and thus becomes *xur*
  - The sine gives the vertical value or *y*, and thus becomes *yur*
  - By definition xur^2 + yur^2 = 1
- The constant π derives from the historical interest in the diameter of a circle for practical building with trees, while current theory shows the importance of the radius. Then Θ = 2 π is the more relevant parameter. This is written as capital theta but pronounced as "archi" (from Archimedes).
- Thus xur[α] = cos[α Θ] and yur[α] = sin[α Θ].

For writing this article, I wondered whether I should use xur and yur here too, rather than cosine and sine. However, the re-engineering of mathematics education differs from the re-engineering of voting theory. It remains useful to explain to voting researchers that the use of cosine and sine is not as complicated at they might think.

[30] Created in: http://fooplot.com

Figure 10. Taken from Colignatus (2009, 2015:65), with its labels Figure 4 and 5

Figure 4. Angular circle ($r = \Theta^H$), unit circle (r = 1), x = Xur and y = Yur

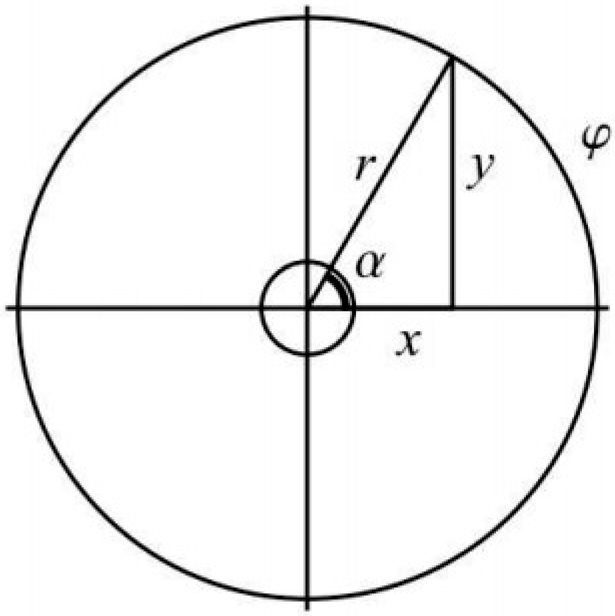
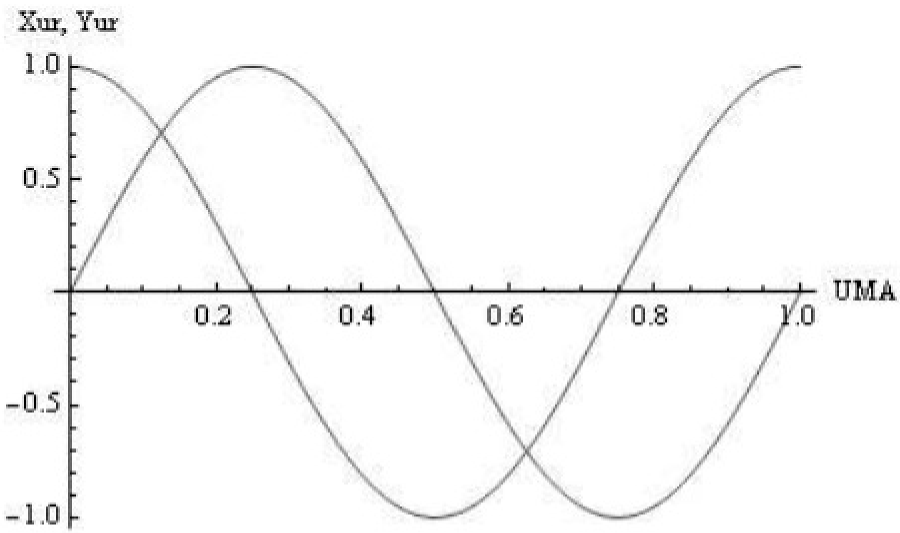


Figure 5. The functional graphs of Xur and Yur



16. Appendix E. The average

The unweighed (RTO) regression z / w = b 1 gives b = Average[z / w]. This can also be seen as the result of regressing z = b w with weights 1 / w, that favour the small parties. [31]

Subsequently, one considers a symmetric expression like A = Abs[Average[z / w] – Average[w^+ / z^+]] in which the division by 0 in the latter denominator is simply removed as a case. The measure looks interesting, as it also has an appeal of junior highschool simplicity.

[31] https://stats.stackexchange.com/questions/54794/regression-through-the-origin

If $w = z$ (and no wasted vote) then $A = 0$.

However, it is not clear why a disproportionality above 1 could compensate a disproportionality below 1, which is what the average does.

The wasted vote causes a major problem too. For votes {99, 1} and seats {100, 0}, the measure becomes Abs[100 / 99 / 2 – 99 / 100 / 1] = 0.485 because the first average divides by length 2 and the second by length 1. We may consider multiplying *A* by 2.

The greatest disproportionality would arise when one party gets all seats. Consider two parties, a range of votes of {*x*, 100 – *x*} and all seats apportioned as {100, 0}. Then the first average gives 100 / *x* / 2 because there is vector of length 2 with one zero. The second average neglects the 0 and divides by 1. Thus:

$$A[\{x, 100 - x\}, \{100, 0\}] = \text{Abs}[100 / x / 2 – x / 100 / 1] = \text{Abs}[50 / x – x / 100]$$

This is large for a small vote (*x* = 1%), becomes zero for *x* 50 Sqrt[2] and, as said, has a value close to ½ for *x* = 100.

It is not clear why the zero for *x* = 50 Sqrt[2] would be the perfect proportional outcome while it isn't.

It appears to be rather complicated to turn *A* or a variant of it into a disproportionality measure. Since there is already a clear measure based upon the angle between the vectors, there doesn't seem to be a need to pursue this line of reasoning on the average.

17. Appendix F. Potential use for apportionment, Holland 2017

Table 14 gives the potential use of SDD for apportionment, applied to the Dutch election for the 2nd Chamber of Parliament in 2017. The scores of the various measures for disproportionality have already been given in Table 4, and see the discussion there. The present table is mainly relevant for the effect on individual parties. The wasted vote will be represented by 3 empty seats or a qualified majority. Large parties lose seats to some smaller ones that were disadvantaged by the use of the method of D'Hondt. Thus this doesn't mean yet that SDD would be a useful method of apportionment, because the same effect might also be gotten by deciding to allow empty seats and using Largest Remainder. Perhaps SDD only helps to highlight the choice.

Table 14. Potential use of SDD for apportionment, example Holland 2017

Comparison								
	Votes	Seats	SDD-Seats	Difference	%Votes	%Seats	%SDD-S	Difference
	10,563,456	150	150	0	100	100	100	0.0
Party								
Other	208,742	0	3	3	2.0	0.0	2.0	2.0
VVD	2,238,351	33	32	-1	21.2	22.0	21.3	-0.7
PVV	1,372,941	20	19	-1	13.0	13.3	12.7	-0.7
CDA	1,301,796	19	18	-1	12.3	12.7	12.0	-0.7
D66	1,285,819	19	18	-1	12.2	12.7	12.0	-0.7
GL	959,600	14	14		9.1	9.3	9.3	
SP	955,633	14	14		9.0	9.3	9.3	
PvdA	599,699	9	8	-1	5.7	6.0	5.3	-0.7
CU	356,271	5	5		3.4	3.3	3.3	
PvdD	335,214	5	5		3.2	3.3	3.3	
50Plus	327,131	4	5	1	3.1	2.7	3.3	0.7
SGP	218,950	3	3		2.1	2.0	2.0	
DENK	216,147	3	3		2.0	2.0	2.0	
FvD	187,162	2	3	1	1.8	1.3	2.0	0.7

18. References

Colignatus is the name in science of Thomas Cool, econometrician and teacher of mathematics, Scheveningen, Holland, http://econpapers.repec.org/RAS/pco170.htm

Belov, D. and R.D. Armstrong (2011), "Distributions of the Kullback-Leibler divergence with applications", J.Math. Stat. Psychol., May, 64(Pt 2):291-309, https://www.ncbi.nlm.nih.gov/pubmed/21492134, res. report: https://www.lsac.org/docs/default-source/research-(lsac-resources)/rr-09-02.pdf

Caulfield, M.J. (2010), "Apportioning Representatives in the United States Congress", MAA Convergence, https://www.maa.org/press/periodicals/convergence/apportioning-representatives-in-the-united-states-congress-introduction

Colignatus, Th. (2006), “On the sample distribution of the adjusted coefficient of determination (R2Adj) in OLS”, http://library.wolfram.com/infocenter/MathSource/6269/

Colignatus, Th. (2007), "Correlation and regression in contingency tables. A measure of association or correlation in nominal data (contingency tables), using determinants", https://mpra.ub.uni-muenchen.de/3660/

Colignatus, Th. (2009, 2015), "Elegance with Substance", https://zenodo.org/record/291974

Colignatus, Th. (2010), "Single vote multiple seats elections. Didactics of district versus proportional representation, using the examples of the United Kingdom and The Netherlands", https://mpra.ub.uni-muenchen.de/22782/

Colignatus, Th. (2011), "Conquest of the Plane", https://zenodo.org/record/291972

Colignatus, Th. 2014), "Voting Theory for Democracy", Thomas Cool Consultancy & Econometrics, https://zenodo.org/record/291985

Colignatus, Th. (2017a), "The performance of four possible rules for selecting the Prime Minister after the Dutch Parliamentary elections of March 2017", https://mpra.ub.uni-muenchen.de/77616/

Colignatus, Th. (2017b), "Two conditions for the application of Lorenz curve and Gini coefficient to voting and allocated seats", https://mpra.ub.uni-muenchen.de/80297/

Dongen, S. van & A.J. Enright (2012), "Metric distances derived from cosine similarity and Pearson and Spearman correlations", https://arxiv.org/abs/1208.3145

Dumont, T. & J.-F. Caulier (2003), "The "effective number of relevant parties": How voting power improves Laakso-Taagepera's index", http://centres.fusl.ac.be/CEREC/document/2003/cerec2003_7.pdf

Draper, N.R. & Y. Yang (1997), "Generalization of the geometric mean functional relationship", Computational Statistics & Data Analysis, Volume 23, Issue 3, 9 January 1997, Pages 355-372; Preprint 1995 Technical report no 943, http://www.stat.wisc.edu/node/1470

Eisenhauer, J.G. (2003), "Regression through the origin", Teaching Statistics. Volume 25, Number 3, Autumn 2003, p76-80

Erb, I. and C. Notredame (2016), "How should we measure proportionality on relative gene expression data?", Theory Biosci. (2016) 135:21–36, https://link.springer.com/content/pdf/10.1007%2Fs12064-015-0220-8.pdf or https://link.springer.com/article/10.1007%2Fs12064-015-0220-8

Gallagher, M. (1991), "Proportionality, disproportionality and electoral systems", Electoral Studies, 10:1, 33-51, https://www.tcd.ie/Political_Science/draft/staff/michael_gallagher/ElectoralStudies1991.pdf

Gallagher, M. (2007), Electoral Systems Web Site, Dublin, Trinity College, http://www.tcd.ie/Political_Science/Staff/Michael.Gallagher/ElSystems/index.php

Gallagher, M. (2017), "Election indices dataset", accessed at 2017-07-31, http://www.tcd.ie/Political_Science/staff/michael_gallagher/ElSystems/Docts/ElectionIndices.pdf

Goldenberg, J. & S.D. Fisher (2017), "The Sainte-Laguë index of disproportionality and Dalton’s principle of transfers", Accepted for publication in Party Politics, March 2017,

Hill, I.D. (1997), "Measuring proportionality", Voting matters, 8, p7-8, http://www.mcdougall.org.uk/VM/VOL1/ISSUE01-23.pdf

Johnston, J. (1972), "Econometric methods", 2nd edition, McGraw-Hill

Karpov, A. (2008), "Measurement of disproportionality in proportional representation systems", Mathematical and Computer Modelling, Volume 48, Issues 9–10, November 2008, Pages 1421-1438, http://www.sciencedirect.com/science/article/pii/S0895717708001933

Kestelman, T. (2005), "Apportionment and Proportionality: A Measured View", Voting Matters, 20, p 12-22, http://www.votingmatters.org.uk/ISSUE20/I20P4.PDF

Koppel, M, and A. Diskin (2009), "Measuring disproportionality, volatility and malapportionment: axiomatization and solutions", Social Choice and Welfare, August, 33:281, https://www.researchgate.net/publication/225444815_Measuring_disproportionality_volatility_and_malapportionment_Axiomatization_and_solutions

Laakso, M. (1980), "Electoral Justice as a criterion for different systems of proportional representation", Scandinavian Political Studies, Bind 3 (New Series) (1980) 3, https://tidsskrift.dk/scandinavian_political_studies/article/view/32355/30159

Laakso, M., & R. Taagepera (2007), "Proportional representation in Scandinavia: Implications for Finland", Scandinavian Political Studies, https://www.researchgate.net/publication/230000636_Proportional_Representation_in_Scandinavia_Implications_for_Finland

Leznik, M. & C. Tofallis (2005), "Estimating Invariant Principal Components Using Diagonal Regression", http://researchprofiles.herts.ac.uk/portal/en/publications/estimating-invariant-principal-components-using-diagonal-regression(d3379080-bda2-4d76-ab8d-b557c7fabea1).html

Lovell, D. V. Pawlowsky-Glahn, J.J. Egozcue, S. Marguerat, J. Bähler (2015), "Proportionality: A Valid Alternative to Correlation for Relative Data", PLOS Computational Biology, http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004075

Malkevitch, J. (2002), "Apportionment", AMS Feature Column, http://www.ams.org/samplings/feature-column/fcarc-apportion1

Malkevitch, J. (2017), "Pairwise Equity in Apportionment (2017)", https://www.york.cuny.edu/~malk/gametheory/tc-2017-pairwise-apportion.html

NCSS Statistical Software (undated), "Lin's concordance correlation coefficient", Chapter 301 of the documentation, https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Lins_Concordance_Correlation_Coefficient.pdf

Pawlowsky-Glahn, V., J.J. Egozcue, R. Meziat (2007), "The statistical analysis of compositional data: The Aitchison geometry", https://laboratoriomatematicas.uniandes.edu.co/cursocoda/04-Vera-geometry.pdf

Pukelsheim, F. (2014), "Proportional representation. Apportionment methods and their applications", Springer,

Renwick, A. (2015), "Electoral Disproportionality: What Is It and How Should We Measure It?", http://blogs.reading.ac.uk/readingpolitics/2015/06/29/electoral-disproportionality-what-is-it-and-how-should-we-measure-it/

Samuelson, P. (1942), " A Note on Alternative Regressions", Econometrica, Vol. 10, No. 1 (Jan., 1942), pp. 80-83, : http://www.jstor.org/stable/1907024

Taagepera, R. and M.S. Shugart (1989), "Seats and Votes", Yale

Taagepera, R. and M. Laakso (2006), "Proportionality Profiles of West European Electoral Systems", European Journal of Political Research 8(4):423 - 446 · May 2006, https://www.researchgate.net/publication/230041613_Proportionality_Profiles_of_West_European_Electoral_Systems

Taagepera, R. and B. Grofman (2003), "Mapping the indices of seats-votes disproportionality and inter-election volatility", Party Politics, 9(6), p659-677, http://escholarship.org/uc/item/0m9912ff#page-1

Tofallis, C. (2000), "Multiple Neutral Regression", Operational Research Paper 14, UHBS 2000:13, http://uhra.herts.ac.uk/bitstream/handle/2299/689/S7.pdf?sequence=1

Quinn, T. (2017), "An introduction to proportionality", https://cran.r-project.org/web/packages/propr/vignettes/a_introduction.html

Zand, M.S., J. Wang, S. Hulchey (2015), "Graphical Representation of Proximity Measures for Multidimensional Data. Classical and Metric Multidimensional Scaling", The *Mathematica* Journal 17, http://www.mathematica-journal.com/2015/09/graphical-representation-of-proximity-measures-for-multidimensional-data/