Working Papers in Statistics
No 2015:12

Department of Statistics
School of Economics and Management
Lund University

Are the Sweden Democrats really Sweden's largest party? A maximum likelihood ratio test on the simplex

JAKOB BERGMAN, LUND UNIVERSITY

BJÖRN HOLMQUIST, LUND UNIVERSITY

Are the Sweden Democrats really Sweden's largest party? A maximum likelihood ratio test on the simplex

Jakob Bergman†

Björn Holmquist

Lund University, Lund, Sweden

Summary. In August 2015 a Swedish newspaper claimed that the Sweden Democrats were the largest political party in Sweden based on the results of single poll. We ask ourselves if this is a correct conclusion, considering the fact that the three largest parties in the poll were of roughly the same size. We analyse the parameter space and identify the subspace where the Sweden Democrats are the largest party. Using this we construct a maximum likelihood ratio test and derive its distribution under the null hypothesis. We finally apply our test to the data and obtaining a p-value between 0.09 and 0.14 are able to refute the claim in the newspaper. Based on the available data one cannot draw the conclusion that the Sweden Democrats are the largest party in Sweden.

Keywords: Isometric logratio transformation; Largest party; Maximum likelihood ratio test; Political polls; Polls, Sweden; Simplex

1. Introduction

On 20 August 2015, the Swedish newspaper *Metro* ran the headline 'Now the Sweden Democrats are Sweden's largest party' (our translation) across its front page (Wallroth, 2015). From a journalistic point of view the headline is not surprising: 10 years ago the nationalistic party the Sweden Democrats (SD) had a voter share of 1–2% and was hardly ever reported in the polls, and now there was a poll that gave the party the largest voter share of any party. A remarkable change indeed. However, from a statistical point of view the headline raised a question: how can we know if SD really are the largest party in the electorate? Any introductory text book in statistics will tell you how to test if a proportion is greater than a specified value in a binomial situation. But in this case there is no specified value to test, and furthermore, Sweden has a multiparty system with 8–10 competing parties, so this is a multinomial situation. How can we, given a vector of observed frequencies, test if a specific share is greater than all the others?

†*Address for correspondence:* Jakob Bergman, Lund University, Dept. of Statistics, Box 743, SE-220 07 Lund, Sweden

E-mail: jakob.bergman@stat.lu.se

One immediate approach would be to perform pairwise tests of the specific share against each of the others. However, to attain an overall level of significance, these tests need to be adjusted, e.g. with a Bonferroni correction. Apart from the general lack of elegance of such an approach, the procedure becomes less attractive when the number of parties increases; in a ten-party situation nine tests would be needed and for each test the significance level would have to be a mere 0.0055 for the overall significance level to be 0.05. A more serious objection is that such procedures do not incorporate the implicit structure of the observed shares or frequencies; due to the fact that they need to sum to 1 or n, respectively, they are not independent but negatively correlated.

Instead of multiple tests, we would like one single test. We propose a maximum likelihood ratio test utilising the inherent properties of shares (proportions) to test the hypothesis. In Section 2 we introduce some notation, formalise the problem and discuss the properties of the parameter space, in Section 3 we derive the test and its properties. We apply the test to the newspaper article above in Section 4.

2. Voter shares and the simplex

Let $\boldsymbol{p} = [p_j]$ denote the vector of voter shares of the D parties in the electorate. (If there is a multitude of very small parties, the Dth share can represent the sum all small parties.) Since $\boldsymbol{p}$ is non-negative and must sum to 1, the parameter space of $\boldsymbol{p}$ is the D-part simplex $\mathbb{S}^D$. Given a simple random sample of n respondents, the number of voters for each party $\boldsymbol{X}$ is a multinomial‡ random variable with parameter $\boldsymbol{p}$.

The statement that the ith share p_i is the greatest of the D shares is a relative statement, which, however, has absolute implications: a necessary condition is that $p_i > 1/D$ and a sufficient condition is that $p_i > 1/2$ (see Appendix for proofs). We believe though that it is easier to consider the entire parameter space than to try to find explicit expressions for p_i. This means testing the hypotheses

$$\begin{aligned} H_0 &: \boldsymbol{p} \in \mathbb{S}^D - \omega_i \\ H_1 &: \boldsymbol{p} \in \omega_i \end{aligned} \tag{1}$$

where ω_i is the subspace of $\mathbb{S}^D$ where the ith part is the greatest. The boundary between the two subspaces is the line, plane etc. where $p_i = p_j$ for at least one $j \neq i$. As an illustration, the parameter space $\mathbb{S}^3$ is depicted in Fig. 1(a) as a ternary diagram.

However, the simplex can pose practical problems due to the constraints on the parameters. Aitchison (1982) introduced the logratio transformations to resolve some of these issues. One popular choice of such transformation is the isometric logratio (ILR) transformation (Egozcue et al., 2003). It resolves the summation constraint of the simplex and transforms the problem to the real space $\mathbb{R}^{D-1}$.

‡$\boldsymbol{X}$ is of course actually hypergeometrically distributed, but we will assume that the population is large enough for the multinomial distribution to be an acceptable approximation.

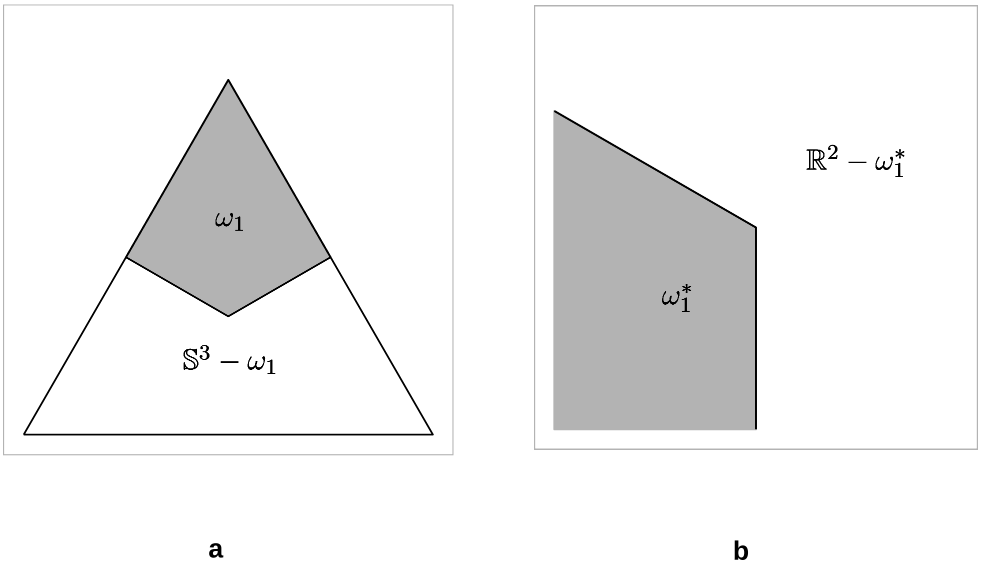

Figure 1. The parameter space $\mathbb{S}^3$ is shown in (a) partitioned into the subspace ω_1, where p_1 is the largest part, and $\mathbb{S}^3 - \omega_1$, where p_1 is not the largest part. The top vertex corresponds to $\boldsymbol{p} = (1, 0, 0)'$, the bottom left to $\boldsymbol{p} = (0, 1, 0)'$, and the bottom right to $\boldsymbol{p} = (0, 0, 1)'$. The boundary between ω_1 and $\mathbb{S}^3 - \omega_1$, is the line from $\boldsymbol{p} = (1/2, 1/2, 0)'$, via $\boldsymbol{p} = (1/3, 1/3, 1/3)'$ to $\boldsymbol{p} = (1/2, 0, 1/2)'$. In (b) the corresponding parameter space in $\mathbb{R}^2$, partitioned into $\omega_1^* = \mathrm{ilr}(\omega_1)$ and $\mathbb{R}^2 - \omega_1^*$, is shown.

As an illustration, the subspaces in $\mathbb{R}^2$ corresponding to $\mathbb{S}^3 - \omega_1$ and ω_1 are depicted in Fig. 1(b). There are many different conceivable ILR transformations, one example is the vector $\boldsymbol{y} = [y_j]$ where

$$y_j = \frac{1}{\sqrt{j(j+1)}} \log \frac{\prod_{k=1}^{j} p_k}{p_{j+1}^{j}}, \qquad j = 1, \ldots, D-1. \tag{2}$$

3. A maximum likelihood ratio test

We propose that (1) is tested using a maximum likelihood ratio (MLR) test. This means finding the maximum value of the likelihood in the restricted parameter space under H_0 and comparing this with the maximum value if the parameter space is not restricted (H_1). As the sample consists of only one observation, the likelihood function equals the probability function:

$$L(\boldsymbol{p}|\boldsymbol{x}) = \frac{n!}{x_1! \cdots x_D!} p_1^{x_1} \cdots p_D^{x_D} \tag{3}$$

The maximum in the unrestricted space is simply the value of the likelihood of the ML estimate $\hat{\boldsymbol{p}} = \boldsymbol{x}/n$. In the restricted parameter space under H_0, the ML estimate is $\boldsymbol{p}^* = \arg\max_{\{\boldsymbol{p} \in \mathbb{S}^D - \omega_i\}} L(\boldsymbol{p}|\boldsymbol{x})$. Assuming that x_i, the part corresponding to p_i, is the largest in the observed vector $\boldsymbol{x}$, the restricted estimate $\boldsymbol{p}^*$ will be a point on the boundary of $\mathbb{S}^D - \omega_i$. This means maximising (3) over $\boldsymbol{p}$ subject to

(a) $p_i \leq p_j$, for all $j \neq i$

(b) $p_j > 0$, for $j = 1, \ldots, D$

(c) $p_1 + \cdots + p_D = 1$

Normally, $\boldsymbol{p}^*$ will have to be estimated numerically. The optimisation is simplified if the last two constraints are removed by transforming the problem to the real space using an ILR transformation. The first constraint (a) can then be reformulated as a set of linear inequalities $\boldsymbol{u}_i \boldsymbol{y} \leq \boldsymbol{0}$ where the matrix $\boldsymbol{u}_i$ will depend upon the exact choice of ILR-transform. If e.g. (2) is used, $D = 5$, and $i = 1$, then

$$\boldsymbol{u}_1 = \begin{bmatrix} \sqrt{2} & 0 & 0 & 0 \\ 1/\sqrt{2} & \sqrt{3/2} & 0 & 0 \\ 1/\sqrt{2} & 1/\sqrt{6} & \sqrt{4/3} & 0 \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} & \sqrt{5/4} \end{bmatrix}$$

The test statistic is

$$\lambda = -2\big(\log L(\hat{\boldsymbol{p}}) - \log L(\boldsymbol{p}^*)\big) = -2\sum_{j=1}^{D} x_j \log \frac{x_j}{np_j^*}. \tag{4}$$

If H_0 is true and $\boldsymbol{p}$ is on the boundary of $\mathbb{S}^D - \omega_i$, then (4) will be 0 with probability δ equal to the proportion of the probability mass located in $\mathbb{S}^3 - \omega_i$. The exact value of δ depends on $\boldsymbol{p}$, but it will typically be close to $1/2$, unless $\boldsymbol{p}$ is close to a point where $d \geq 3$ parts are equal, in which case δ will be close to $(d-1)/d$. The number of such points increases with the number of parts D and equals $\sum_{k=3}^{D}\binom{D}{k}$. With probability $1-\delta$, (4) will be asymptotically χ^2-distributed with one degree of freedom. Even though we only have one observation of $\boldsymbol{x}$, this observation is based on n respondents, and as long as n is sufficiently large the asymptotic result will hold. The one degree of freedom follows from the fact that under H_1 we estimate $D-1$ parameters freely, but under H_0 we restrict p_i to be equal to one of the other $D-1$ estimated parameters and hence only $D-2$ parameters are estimated freely.

The p-value of the test may be obtained as

$$\frac{1-F(\lambda)}{d},$$

where F is the cumulative density function of the χ^2-distribution with one degree of freedom and d is the number of estimates in $\boldsymbol{p}^*$ equal or almost equal to p_i^* (including p_i^*).

4. Are the Sweden Democrats Sweden's largest party?

The poll that the newspaper *Metro* published was done by YouGov Sweden. In total 1527 people responded to the poll. Nine parties were reported yielding the estimates $\hat{\boldsymbol{p}}$ in Table 1. We note that the three largest parties, the Moderates (M), the Social Democrats (S), and SD, are roughly equal in size: 21–25%. Since the reported shares of the nine parties sum to 0.993, the share of all other parties must sum to 0.007.

Table 1. The reported estimated voter shares $\hat{\boldsymbol{p}}$, the corresponding frequencies $\boldsymbol{x}$ (assuming a simple random sample), and the estimated voter shares $\boldsymbol{p}^*$ under the restriction that SD are not allowed to be the largest party.

Party[a]	M	C	FP	KD	MP	S	V	FI	SD	Other
$\hat{\boldsymbol{p}}$	0.210	0.056	0.044	0.037	0.064	0.234	0.068	0.028	0.252	0.007
$\boldsymbol{x}$	321	85	67	56	98	357	104	43	385	11
$\boldsymbol{p}^*$	0.210	0.055	0.043	0.036	0.064	0.244	0.070	0.029	0.244	0.007

[a] The Moderates (M), the Centre Party (C), the Liberals (FP), the Christian Democrats (KD), the Green Party (MP), the Social Democrats (S), the Left Party (V), the Feminist Party (FI), and the Sweden Democrats (SD).

YouGov used a a self-recruited on-line panel for the poll, i.e. not a random sample from the electorate, and most likely weighted the answers in some intricate way. However, we will assume that the estimates are based on a simple random sample from the electorate. Given a sample of 1527 respondents, the reported shares would correspond to the frequencies $\boldsymbol{x}$ in Table 1.

The ML estimate $\boldsymbol{p}^*$ in the restricted parameter space, i.e. the space where SD are not the largest party, is given in Table 1. We note that the estimated shares of S and SD are equal, as $\boldsymbol{p}^*$ is restricted to $\mathbb{S}^{10} - \omega_{\text{SD}}$ including its boundary. The estimate $\boldsymbol{p}^*$ corresponds to a log likelihood value of $\log L(\boldsymbol{p}^*) = -28.038$, whereas the log likelihood value for the ML estimate in the unrestricted parameter space is $\log L(\hat{\boldsymbol{p}}) = -27.452$. This gives a test statistic of $\lambda = 1.171$, yielding a p-value between $(1-F(1.171))/2 = 0.14$ and $(1-F(1.171))/3 = 0.09$. Based on the data, we cannot draw the conclusion that the Sweden Democrats are the largest party in Sweden and we can refute the claim made by *Metro*.

5. Concluding remarks

We have introduced a novel maximum likelihood ratio test for testing if a specific part or proportion is the largest among D parts. Although the test was conceived with regard to a certain newspaper article, the test is very general and can be applied to any situation, with a finite number of parts, where one wants to test if the observed frequencies support the hypothesis that a specific share of the population is the largest. The test is fairly straightforward and attains a specified size. A small simulation study indicates that the empirical size of the test is reasonably close to the theoretical size; in our simulations the empirical sizes varied from 0.04 to 0.07 compared with the theoretical size 0.05. An important aspect of the test is that it is based on, and respects, the parameter space of the problem.

It remains as future research to develop an exact expression for δ as a function of $\boldsymbol{p}^*$. However, at this point we believe that such expressions would be of more theoretical value than of dramatically improved practical usefulness.

We hope that this test can be of practical use for many, not at least when analysing political poll results. Instead of making unsupported claims, as *Metro*, or conducting a lot of tests, this test provides a simple and stringent way of testing a statement of great interest to many people.

Appendix

THEOREM 1. *If* $\boldsymbol{p} \in \mathbb{S}^D$ *and* p_i *is the largest part, then a necessary condition is that* $p_i > 1/D$.

PROOF. Let p_i be the largest part. If $p_i < 1/D$, then all other parts are also less than $1/D$, and the sum of all parts is then less than 1. If $p_i = 1/D$ then either $p_j < 1/D$ for some $j \neq i$ and the sum of all parts is less than 1, or $p_j = 1/D$ for all $j = 1, \ldots, D$ but then p_i is not the largest part.

THEOREM 2. *If* $\boldsymbol{p} \in \mathbb{S}^D$ *and* p_i *is the largest part, then a sufficient condition is that* $p_i > 1/2$.

PROOF. If p_i is not the largest part and $p_i > 1/2$, then there exists a part $p_j > p_i$ for some $j \neq i$. But then also $p_j > 1/2$ and the sum of the parts is greater than 1.

References

Aitchison, J. (1982) The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **44**, 139–177.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barceló-Vidal, C. (2003) Isometric logratio transformations for compositional data analysis. *Math. Geol.*, **35**, 279–300.

Wallroth, E. (2015) Nu är SD Sveriges största parti. *Metro.* Skåne Edition. 20 August 2015.

http://journals.lub.lu.se/stat

LUND UNIVERSITY
School of Economics and Management

Working Papers in Statistics 2015
LUND UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT
Department of Statistics
Box 743
220 07 Lund, Sweden