Reduced-bias inference for multi-dimensional Rasch models with applications

Ioannis Kosmidis[1]
i.kosmidis@ucl.ac.uk
David Firth[2]
d.firth@warwick.ac.uk
Heather Turner[2]
ht@heatherturner.net

[1] Department of Statistical Science, UCL
[2] Department of Statistics, University of Warwick

IWSM 2013
July 2013, Palermo, Italy

Rasch models

- Independent Bernoulli responses in a subject-item arrangement: Y_{is} is the outcome of the sth subject on the ith item.
- $\pi_{is} = P(Y_{is} = 1)$: the probability that sth subject succeeds on the ith item, $(i = 1, \ldots, I; s = 1, \ldots, S)$.

2PL model

- The 2PL Rasch model:

$$\log \frac{\pi_{is}}{1-\pi_{is}} = \eta_{is} = \alpha_i + \beta_i \gamma_s \quad (i = 1, \ldots, I; s = 1, \ldots, S)\,,$$

- Parameter interpretation:
  - α_i (or $-\alpha_i$): measure of the "ease" (or "difficulty") of the ith item,
  - β_i: a "discrimination" parameter for the ith item,
  - γ_s: the "ability" of the sth subject.

Extensions

- More than one "discrimination" and "ability" dimensions:

$$\log \frac{\pi_{is}}{1-\pi_{is}} = \eta_{is} = \alpha_i + \sum_{j=1}^{m} \beta_{ji}\gamma_{js} \quad (i = 1, \ldots, I; s = 1, \ldots, S) .$$

- $\theta = (\alpha^T, \beta_1^T, \ldots, \beta_m^T, \gamma_1^T, \ldots, \gamma_m^T)^T$.
- The number of parameters is $p = I + m(I + S)$

Scaling of legislators

Data: US House of Representatives 2001:

- 20 roll calls selected by *Americans for Democratic Action* (ADA).

Legislator	Roll call							
	1	2	3	4	5	6	7	8	. . .
Akin	0	0	0	0	0	0	0	0	. . .
Allen	1	1	1	1	1	1	1	1	. . .
Andrews	1	1	1	1	1	1	1	1	. . .
Armey	0	0	0	0	0	0	0	0	. . .
Baca	1	1	1	1	1	1	1	NA	. . .

- the agreement of the votes of 435 legislators to ADAs position was recorded.
- Aim: Place the legislators on a “liberality” scale.

Data kindly supplied by Jan deLeeuw, used in deLeeuw (2006, *CSDA*).

Maximum likelihood estimation - Issues

→ ML estimation is straighforward using generic tools (e.g. gnm uses a quasi Newton-Raphon iteration).

- Useful asymptotic frameworks (e.g. information grows with the number of subjects):
  → Full maximum likelihood generally delivers inconsistent estimates (Andersen, 1980, Chapter 6).
  → Loss in performance of tests, confidence intervals.

- (Partial) Solutions: Integrated likelihoods, modified profile likelihoods
  → can be hard to apply for 2PL and extensions due to nonlinearity.

Maximum likelihood estimation - Issues

- There is positive probability of boundary ML estimates.
  - → Numerical issues in estimation.
  - → Problems with asymptotic inference (e.g. Wald-type).
- Add small constants to the responses and totals (Haldane, 1955, Annals of Human Genetics).
  - → Arbitrariness of the choice of constants
  - → Not generally a good idea (K., 2013, JRSSB).

Bias-reducing adjusted score functions

- K. and Firth (2009, B'ka) : appropriate adjustment $A(\theta)$ to the score vector for getting estimators with smaller asymptotic bias than ML:

$$\nabla_\theta l(\theta) + A(\theta) = 0\,.$$

- Applicable to models where the information increases with the number of observations ($\dim\theta$ is independent of the number of observations).
  $\rightarrow$ Not the case for Rasch models under useful asymptotic frameworks.
  $\rightarrow$ But expect less-biased estimators than ML.

Bias-reducing adjusted score functions

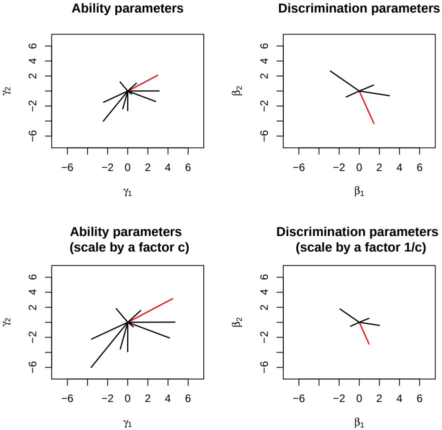
→ In binomial/multinomial response GLMs, the reduced-bias estimates are always finite (Heinze and Schemper 2002, StatMed; K. 2013, JRSSB)

→ Easy implementation through Iterated ML fits on pseudo-data (K. and Firth, 2011, B'ka)

- An identifiable parameterization is necessary.

Identifiability: 2-dimensional model

Over-parameterized version: $\mathrm{logit}(\pi_{is}) = \alpha_i + \beta_{1i}\gamma_{1s} + \beta_{2i}\gamma_{2s}$

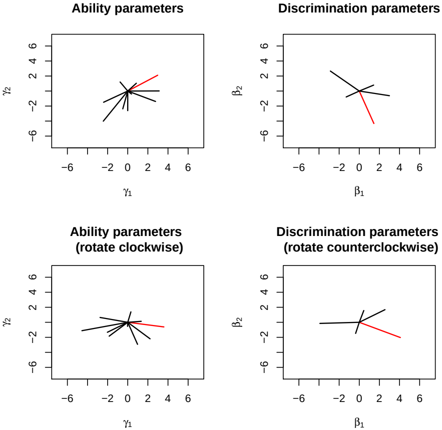


→ Fixed location: α and $(\gamma_{j1}, \ldots, \gamma_{jS})$.

→ Fixed scale: $(\beta_{j1}, \ldots, \beta_{jI})$ and $(\gamma_{j1}, \ldots, \gamma_{jS})$.

Identifiability: 2-dimensional model

Over-parameterized version: $\operatorname{logit}(\pi_{is}) = \alpha_i + \beta_{1i}\gamma_{1s} + \beta_{2i}\gamma_{2s}$



$\rightarrow$ Fixed orientation: (β_{1i}, β_{2i}) and $(\gamma_{1s}, \gamma_{2s})$.

Identifiable parameterization

- $$\log \frac{\pi_{is}}{1-\pi_{is}} = \eta_{is} = \alpha_i + \sum_{j=1}^{m} \beta_{ji}\gamma_{js} \quad (i = 1, \ldots, I; s = 1, \ldots, S)\,.$$

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \ldots & \beta_{1I} \\ \beta_{21} & \beta_{22} & \ldots & \beta_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{m1} & \beta_{m2} & \ldots & \beta_{mI} \end{bmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \ldots & \gamma_{1S} \\ \gamma_{21} & \gamma_{22} & \ldots & \gamma_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \ldots & \gamma_{mS} \end{bmatrix}$$

- A sufficient set of constraints for identifiability:
  - Fix exactly m columns of B.
  - Fix exactly 1 element from each row of Γ.

$\rightarrow$ A total of $m(m+1)$ constraints.

- From the $p = I + m(I + S)$ parameters, only $p_E = I + m(I + S - m - 1)$ are effective.

Adjusted score equations for Rasch models

Adjusted score equations (only p_E effective) (K. and Firth, 2009, B'ka)

$$0 = \sum_{i=1}^{I} \sum_{s=1}^{S} \left(y_{is} + \frac{1}{2} h_{is} + (1 + h_{is}) \pi_{is} + c_{is} v_{is} \right) z_{ist} \quad (t = 1, \ldots, p) \,,$$

where

- $z_{ist} = \partial \eta_{is} / \partial \theta_t$ is the (s, t)th element of the $S \times (2I + S)$ matrix Z_i,
- is the sth diagonal element of $H_i = Z_i F^{-1} Z_i^T \Sigma_i$ ("hat value" for the (i, s)th observation),
- $F = \sum_{i=1}^{T} Z_i^T \Sigma_i Z_i$,
- $\Sigma_i = \mathrm{diag}\left\{v_{i1}, \ldots, v_{iS}\right\}$, $v_{is} = \mathrm{var}(Y_{is}) = \pi_{is}(1 - \pi_{is})$,
- $c_{is} = \sum_{j=1}^{m} \mathrm{AsCov}(\beta_{ji}, \gamma_{js})$ ($\mathrm{AsCov}(\beta_{ji}, \gamma_{js})$ from the appropriate components of F^{-1}).

Comparison with ML equations for Rasch models

Adjusted score equations (only p_E effective) (K. and Firth, 2009, B'ka)

$$0 = \sum_{i=1}^{I} \sum_{s=1}^{S} \left(y_{is} + \frac{1}{2} h_{is} + (1 + h_{is}) \pi_{is} + c_{is} v_{is} \right) z_{ist} \quad (t = 1, \ldots, p) \, ,$$

where

- $z_{ist} = \partial \eta_{is} / \partial \theta_t$ is the (s, t)th element of the $S \times (2I + S)$ matrix Z_i,
- is the sth diagonal element of $H_i = Z_i F^{-1} Z_i^T \Sigma_i$ ("hat value" for the (i, s)th observation),
- $F = \sum_{i=1}^{T} Z_i^T \Sigma_i Z_i$,
- $\Sigma_i = \mathrm{diag}\{v_{i1}, \ldots, v_{iS}\}$, $v_{is} = \mathrm{var}(Y_{is}) = \pi_{is}(1 - \pi_{is})$,
- $c_{is} = \sum_{j=1}^{m} \mathrm{AsCov}(\beta_{ji}, \gamma_{js})$ $(\mathrm{AsCov}(\beta_{ji}, \gamma_{js})$ from the appropriate components of F^{-1}).

Pseudo data

→ If h did not depend on the parameters then the reduced-bias estimator would be formally the ML estimator on Binomial pseudo-data.

Pseudo data	
Responses:	$y^* = y + h/2 + c\pi(1-\pi)$
Totals:	$m^* = 1 + h$

Pseudo data

→ If h did not depend on the parameters then the reduced-bias estimator would be formally the ML estimator on Binomial pseudo-data.

Pseudo data	
Responses:	$y^* = y + h/2 + c\pi 1_{(c>0)}$
Totals:	$m^* = 1 + h + c(\pi - 1_{(c<0)})$

* via algebraic manipulation of the adjusted scores to ensure $0 \leq y^* \leq m^*$. Here, $1_E = 1$ if E holds.

Iterated ML fits on pseudo data

- The adjusted score equations can be solved as follows.

Iterated ML fits on pseudo data

At each iteration

1. Update the values of the pseudo data.
2. Use ML to fit the Rasch model on the current value of the pseudo data.

Repeat until the changes to the estimates are small.

- Ingredients: standard ML software, routines for extracting the hat values and Fisher information.

$\rightarrow$ `gnm` and the methods `hatvalues`, `vcov` for `gnm` objects can do this

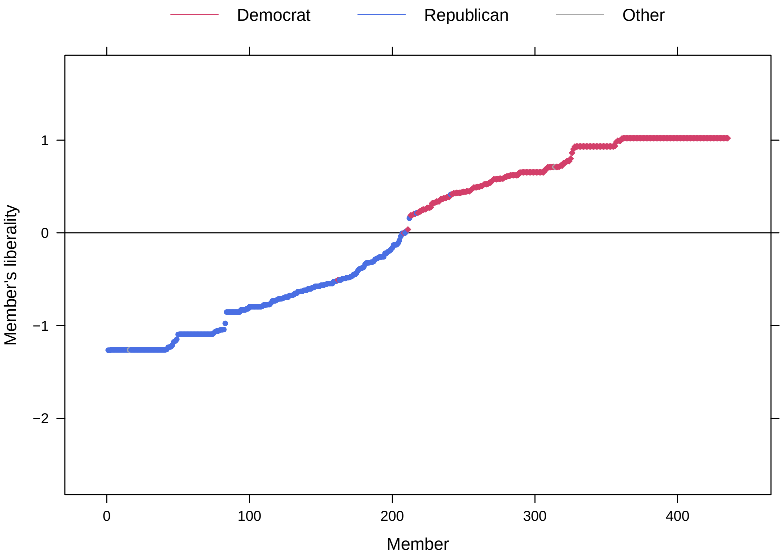
Scaling of legislators

Data: US House of Representatives 2001:

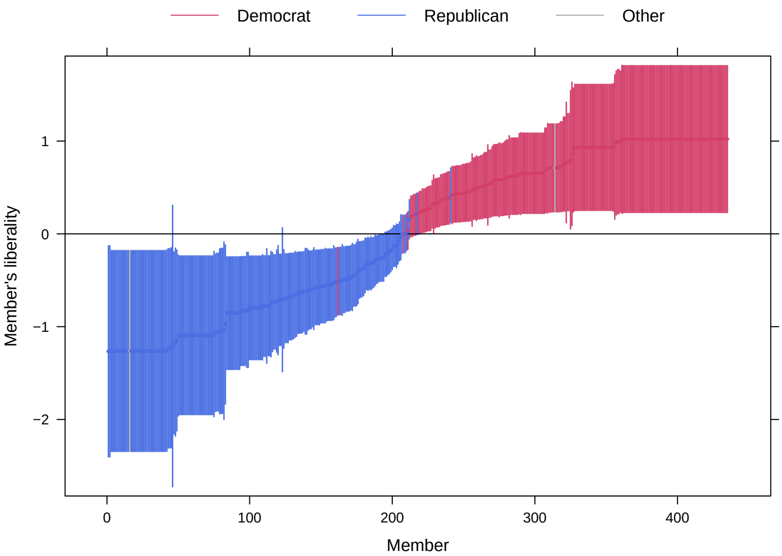
- 20 roll calls selected by *Americans for Democratic Action* (ADA).
- Aim: Place the 435 legislators on a "liberality" scale.

Model	$\dim\theta$	Effective
1-dim	475	473
2-dim	930	924

Results from the one-dimensional model

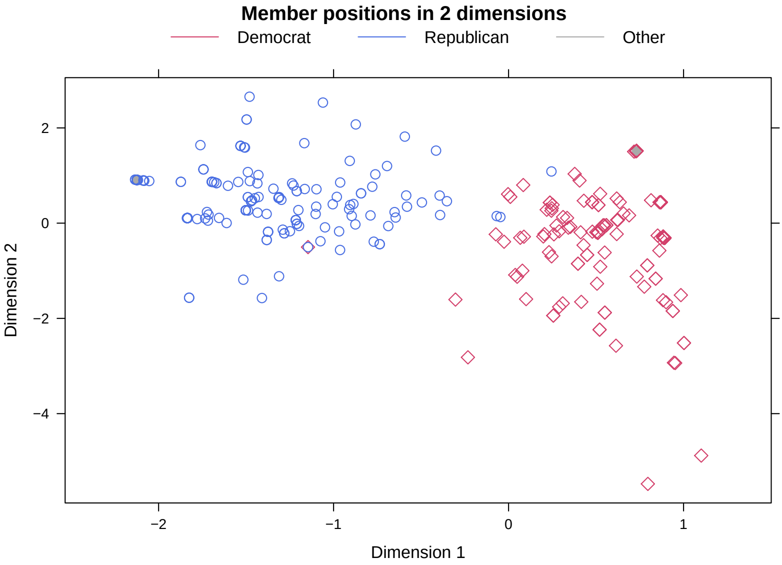


Results from the one-dimensional model

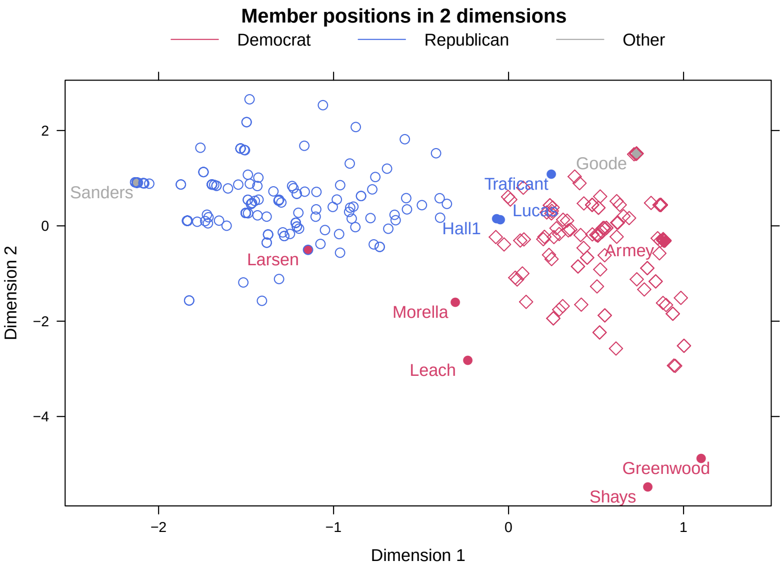


(for details on 'comparison intervals' see Firth and de Menezes, 2004, B'ka.)

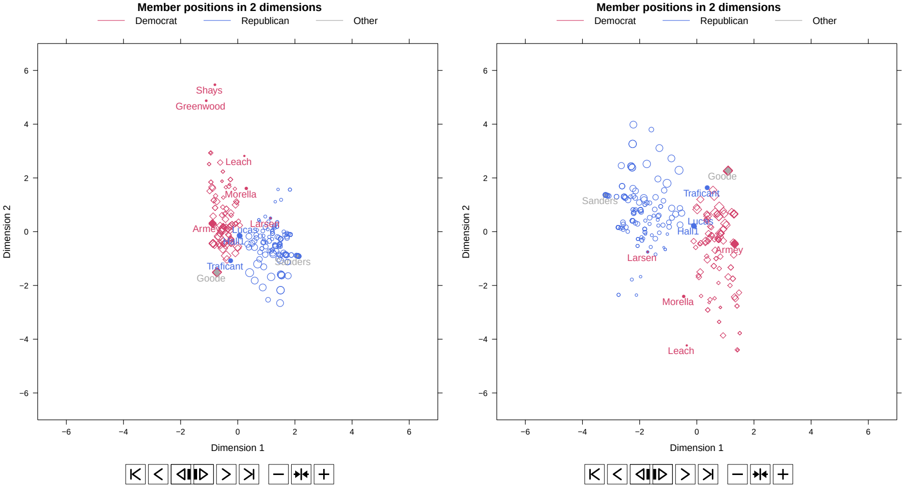
Results from the two-dimensional model



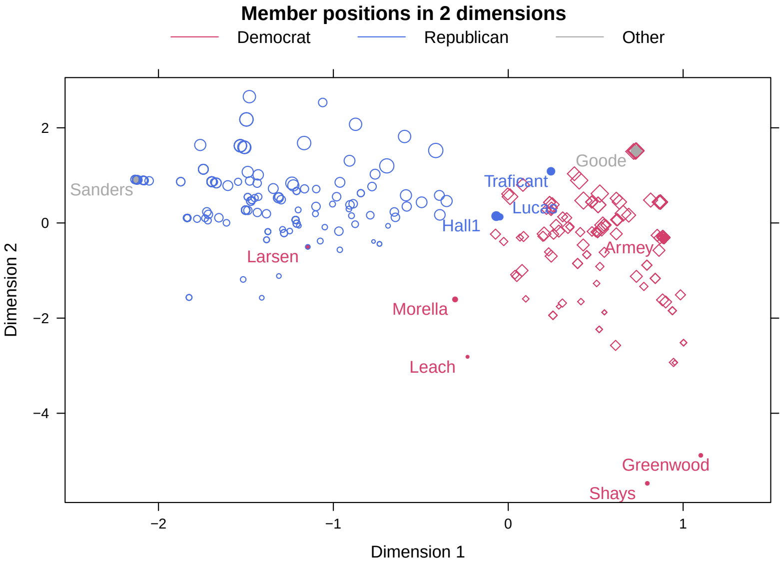
Results from the two-dimensional model



Equivariance under rotation, scale changes and translation



Interpretation of the dimensions of liberality



size: difference “for % on economic matters” − “for % on social matters”

Discussion

→ The method described here yields more sensible results than either *MLE* or *constant* data-adjustment.

→ Computationally convenient.

→ But still it is *inconsistent* (e.g., as the number of items increases).

→ Like the MLE, the resultant estimators are equivariant under the "interesting" transformations (rotation, scale changes, translation). But they are **not** equivariant for general transformations.

- Extensions to time-dependent liberality scales.

Firth, D. and R. X. de Menezes (2004). Quasi-variances. *Biometrika 91*(1), 65–80.

Haldane, J. (1955). The estimation of the logarithm of a ratio of frequencies. *Annals of Human Genetics 20*, 309–311.

Heinze, G. and M. Schemper (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine 21*, 2409–2419.

K., I. (2009). On iterative adjustment of responses for the reduction of bias in binary regression models. Technical Report 09-36, CRiSM working paper series.

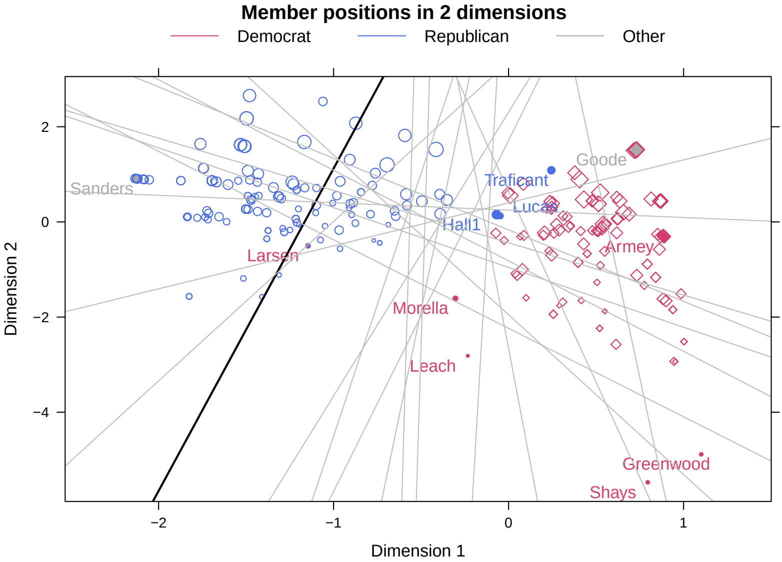
K., I. (2013, March). Improved estimation in cumulative link models. *ArXiv e-prints 1204.0105. To appear in the Journal of the Royal Statistical Society: Series B*.

K., I. and D. Firth (2009). Bias reduction in exponential family nonlinear models. *Biometrika 96*(4), 793–804.

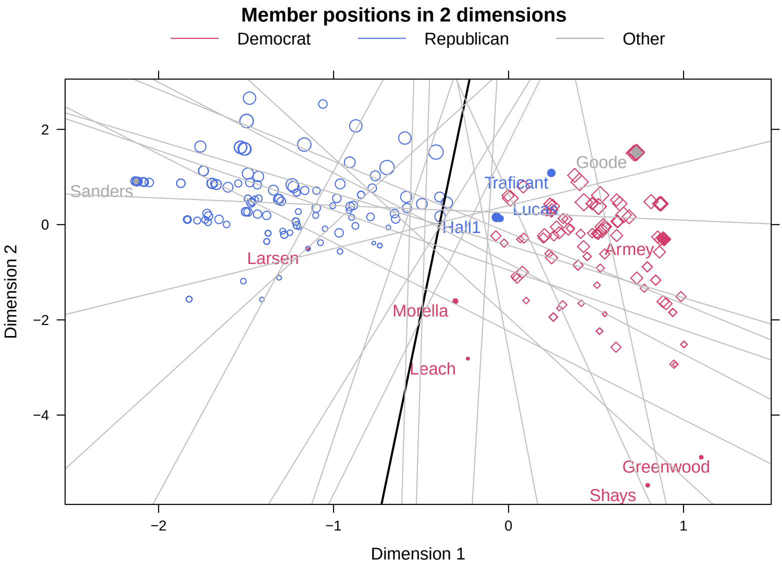
K., I. and D. Firth (2010). A generic algorithm for reducing bias in parametric estimation. *Electronic Journal of Statistics 4*, 1097–1112.

K., I. and D. Firth (2011). Multinomial logit bias reduction via the poisson log-linear model. *Biometrika 98*(3), 755–759.

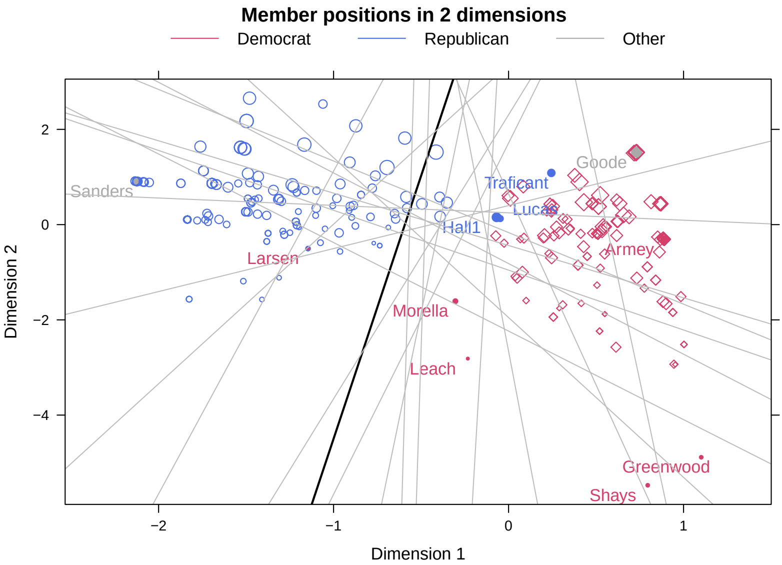
Separation per roll call



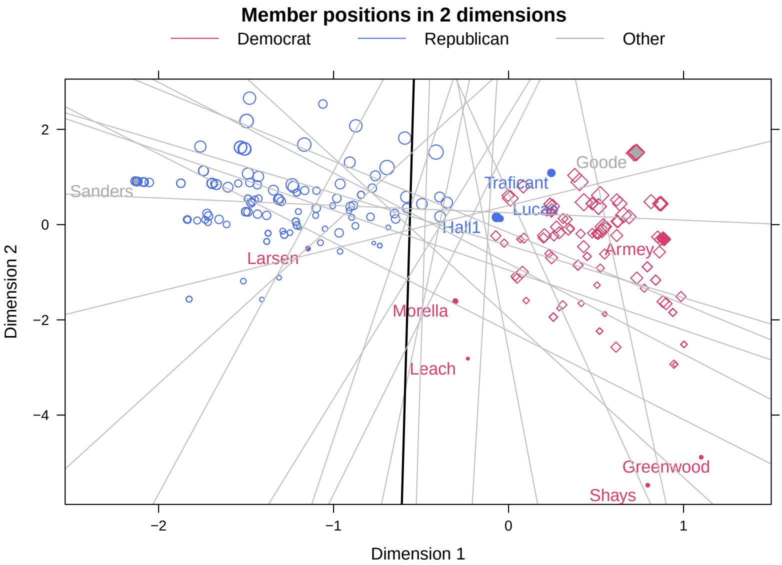
Separation per roll call



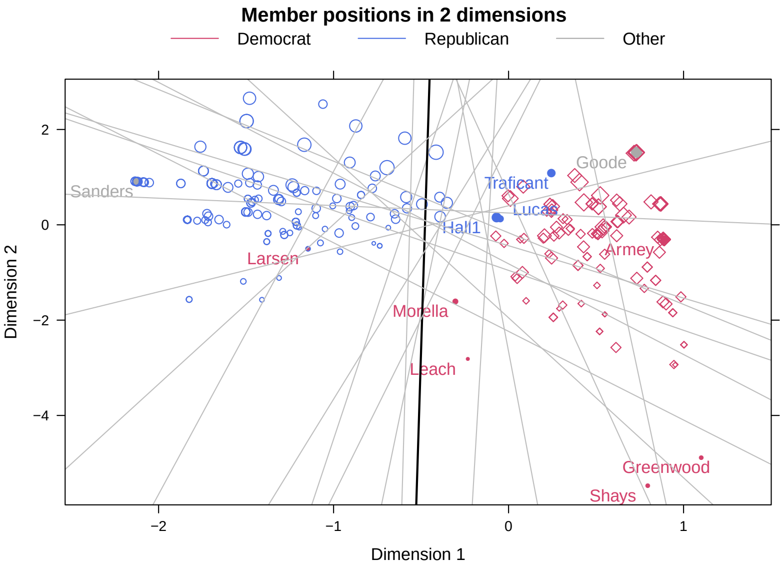
Separation per roll call



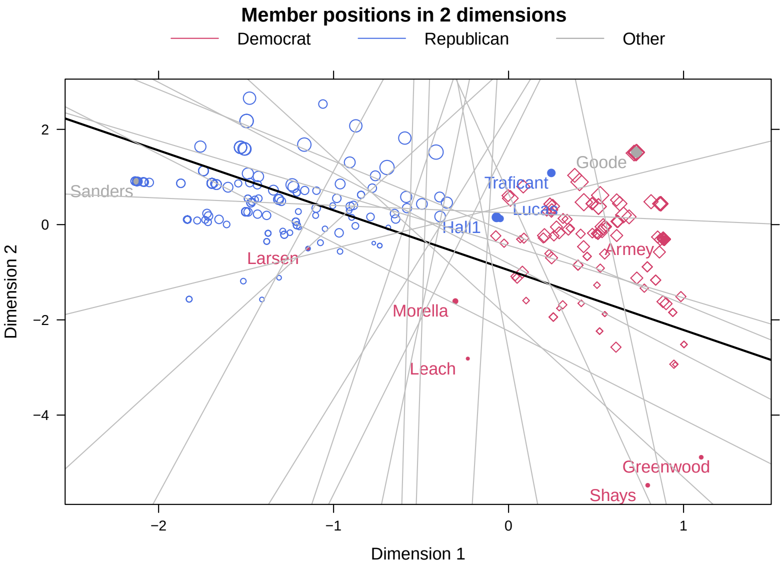
Separation per roll call



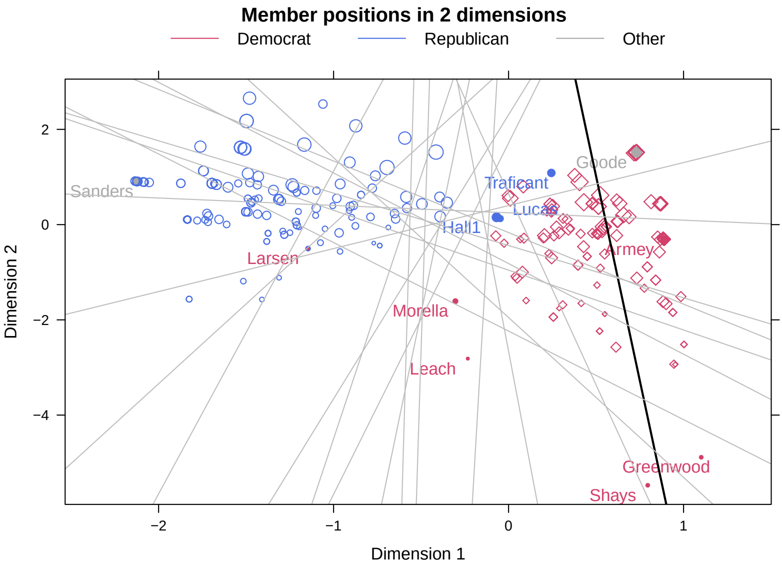
Separation per roll call



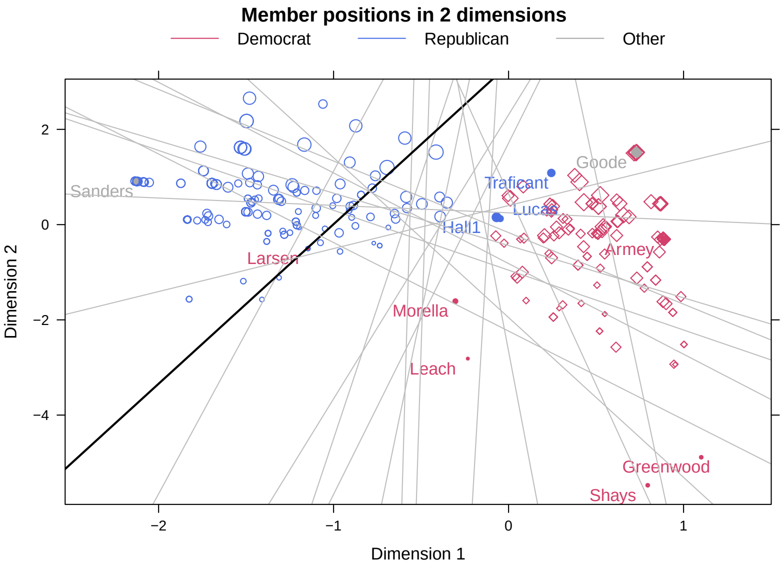
Separation per roll call



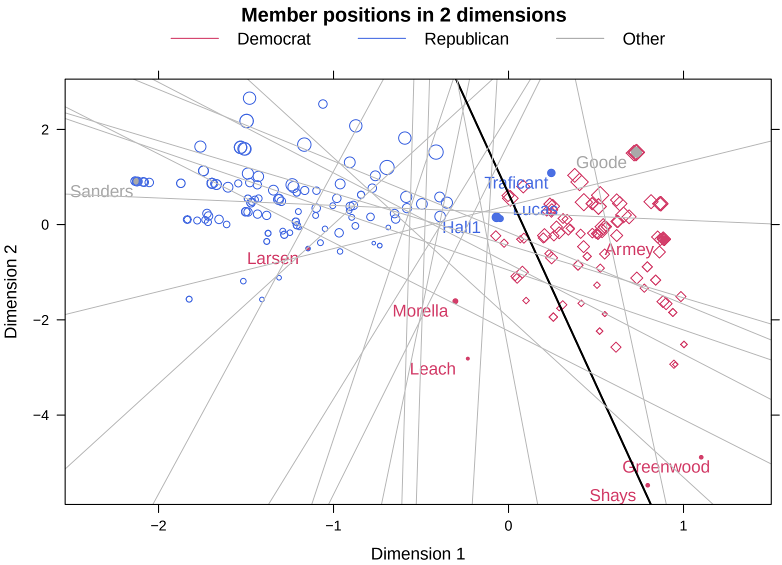
Separation per roll call



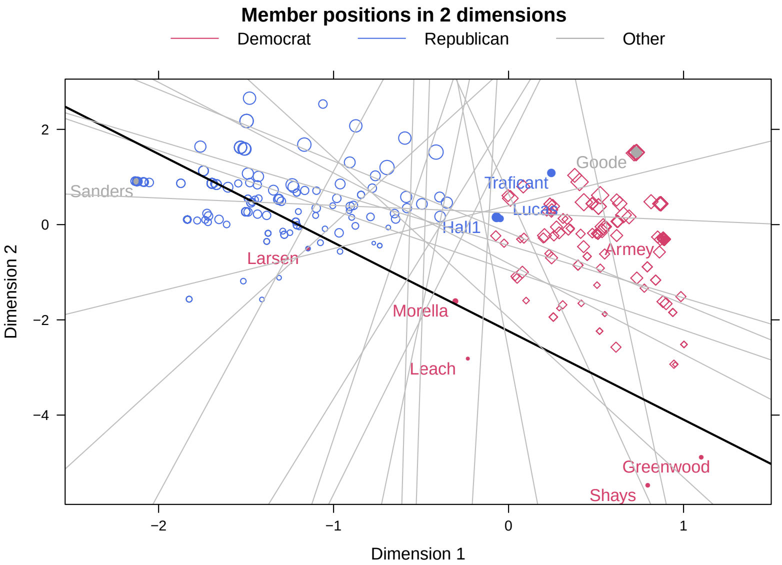
Separation per roll call



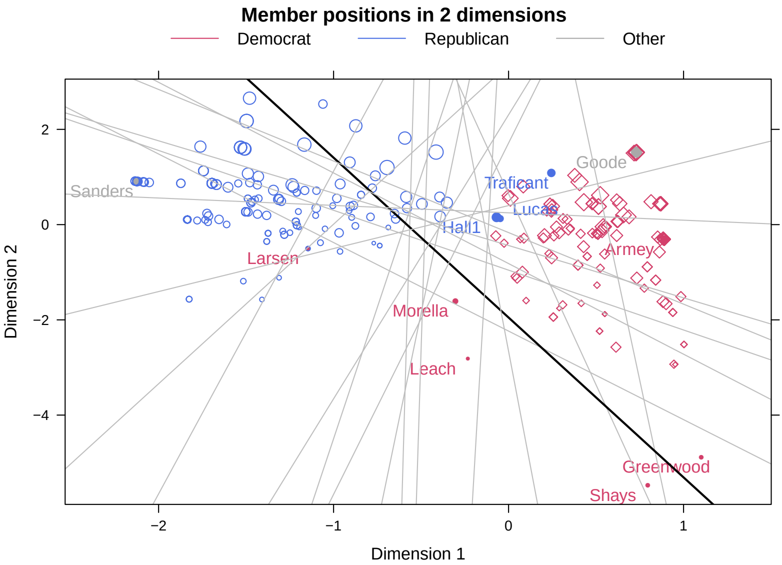
Separation per roll call



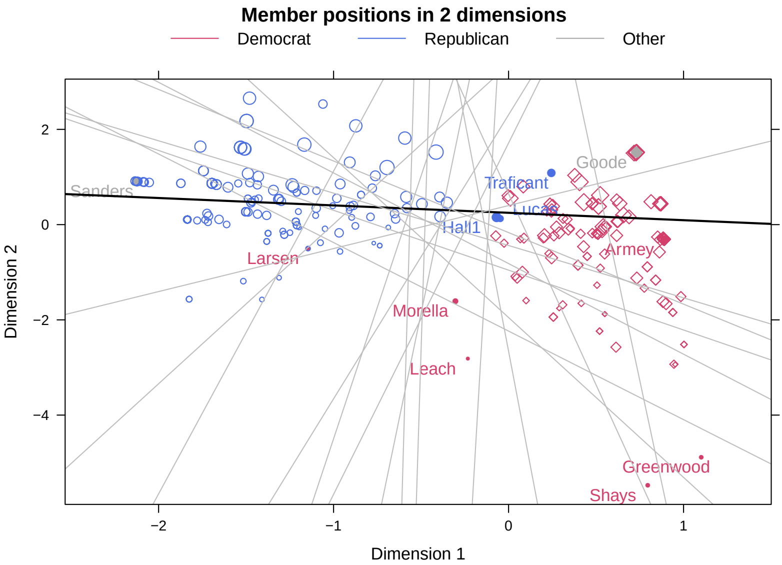
Separation per roll call



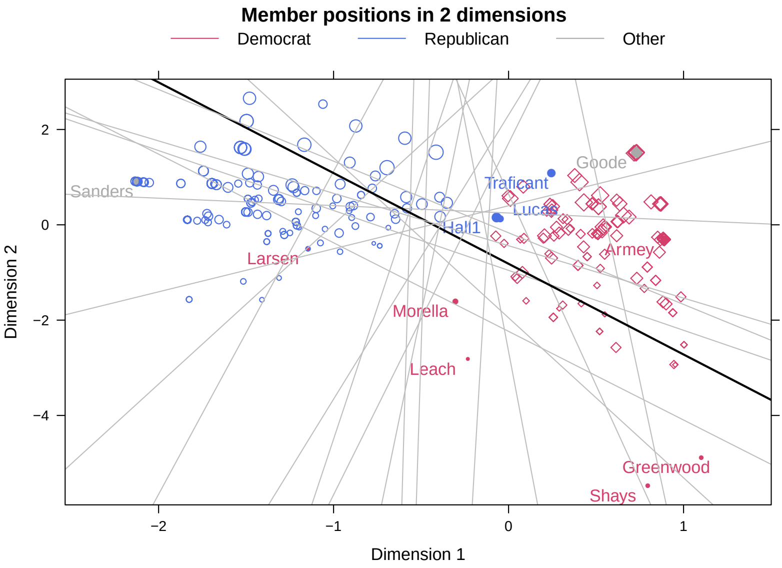
Separation per roll call



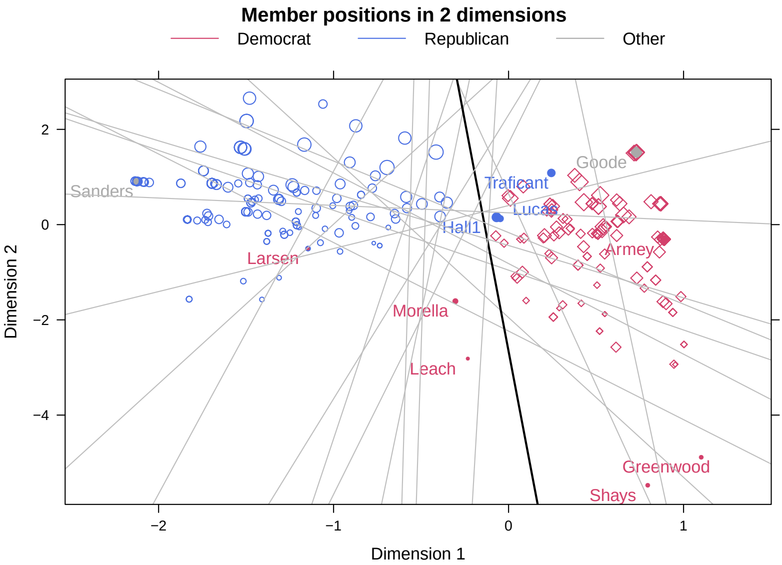
Separation per roll call



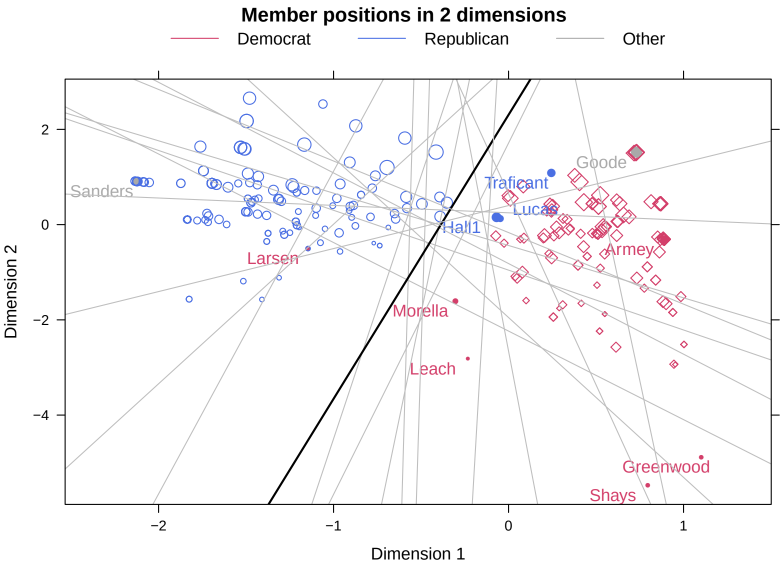
Separation per roll call



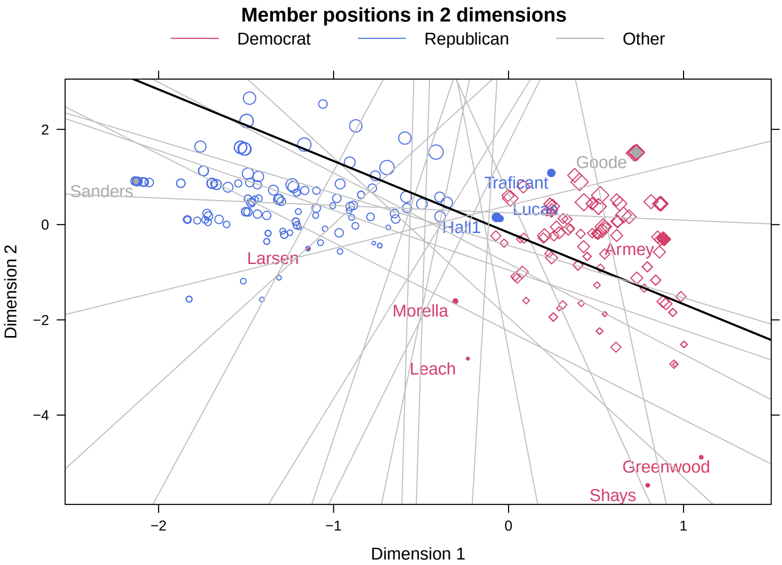
Separation per roll call



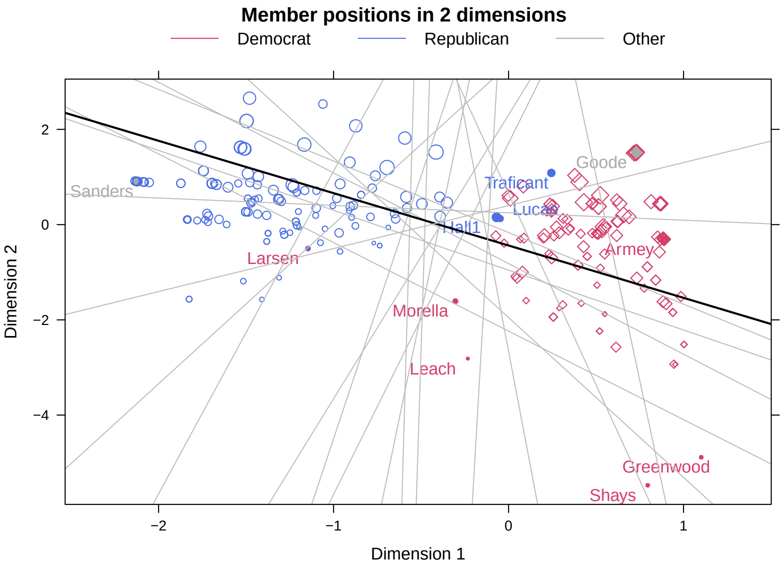
Separation per roll call



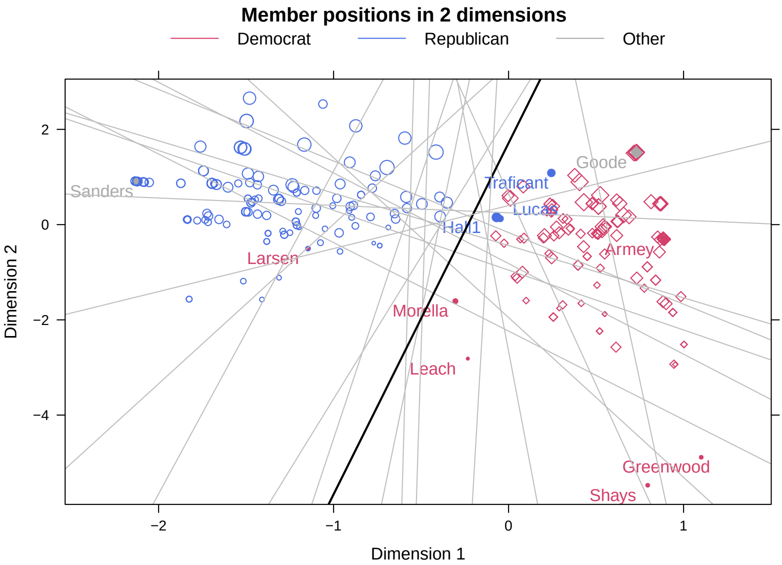
Separation per roll call



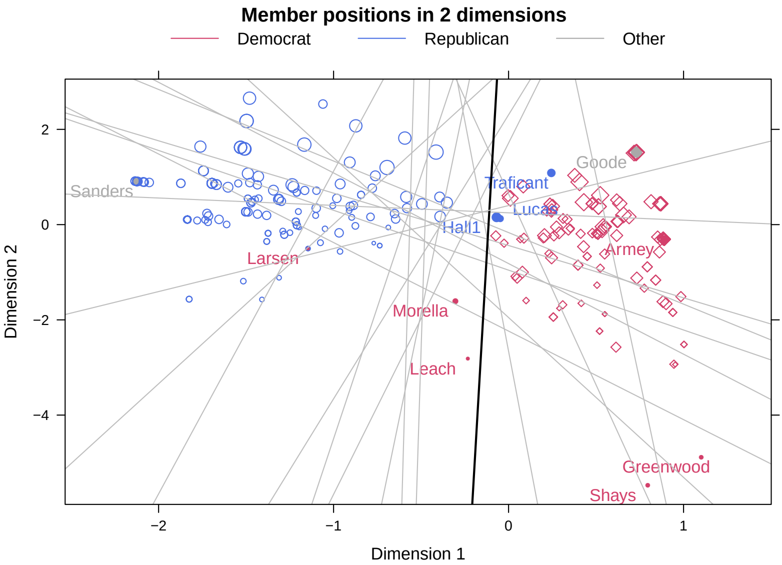
Separation per roll call



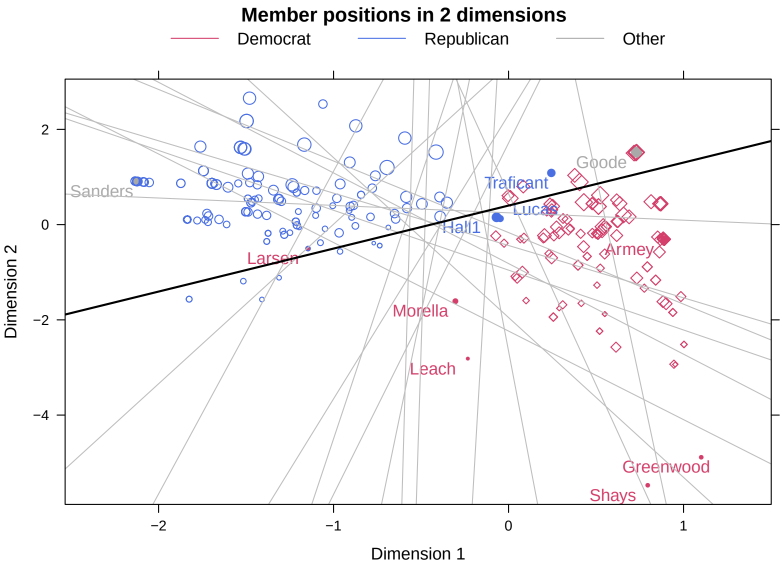
Separation per roll call



Separation per roll call



Separation per roll call



Firth, D. and R. X. de Menezes (2004). Quasi-variances. *Biometrika 91*(1), 65–80.

Haldane, J. (1955). The estimation of the logarithm of a ratio of frequencies. *Annals of Human Genetics 20*, 309–311.

Heinze, G. and M. Schemper (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine 21*, 2409–2419.

K., I. (2009). On iterative adjustment of responses for the reduction of bias in binary regression models. Technical Report 09-36, CRiSM working paper series.

K., I. (2013, March). Improved estimation in cumulative link models. *ArXiv e-prints 1204.0105. To appear in the Journal of the Royal Statistical Society: Series B*.

K., I. and D. Firth (2009). Bias reduction in exponential family nonlinear models. *Biometrika 96*(4), 793–804.

K., I. and D. Firth (2010). A generic algorithm for reducing bias in parametric estimation. *Electronic Journal of Statistics 4*, 1097–1112.

K., I. and D. Firth (2011). Multinomial logit bias reduction via the poisson log-linear model. *Biometrika 98*(3), 755–759.