Title
Principal Component Analysis of Binary Data. Applications to Roll-Call-Analysis

Author
Jan de Leeuw

Publication Date
2011-10-25

PRINCIPAL COMPONENT ANALYSIS OF BINARY DATA. APPLICATIONS TO ROLL-CALL ANALYSIS

JAN DE LEEUW

ABSTRACT. We compute the maximum likelihood estimates of a principal component analysis on the logit or probit scalem using a majorization algorithm that computes a sequence of singular value decompositions. The technique is applied to 2001 house and senate roll call data and compared with other techniques for roll call analysis.

1. INTRODUCTION

Suppose $P = \{p_{ij}\}$ is an $n \times m$ binary data matrix, i.e. a matrix with elements equal to zero or one (or to yes/no, true/false, present/absent, agree/disagree). For the moment we suppose that P is *complete*, the case in which some elements are *missing* is discussed in a later section.

There are many examples of such binary data in the sciences. We give a small sample in the table below, many more could be added.

TABLE 1. Binary data

discipline	**rows**	**columns**
political science	legislators	roll-calls
education	students	test items
systematic zoology	species	characteristics
ecology	plants	transects
archeology	artefacts	graves
sociology	interviewees	questions

Date: August 24, 2003.

2000 *Mathematics Subject Classification.* 62H25,62H30,62P25.

Key words and phrases. Multivariate Analysis, Principal Components, Classification, Applications to social sciences.

In this paper we will concentrate on the analysis of roll call data, but it goes without saying that our results can be applied to the other examples in Table 1 as well.

There are many different techniques that have been used to analyze data of this form. One, important class are *latent structure techniques*, which include latent class analysis, latent trait analysis and various forms of factor analysis for binary data. By recoding the data as a 2^m table, *log-linear* decompositions and approximations of the multivariate distribution become available. There are also various forms of *cluster analysis* which can be applied to binary data, usually by first computing some sort of similarity measure between rows and/or column. And then there are variations of *principal component analysis* for binary data such as multiple correspondence analysis.

We combine ideas of latent structure analysis, more particularly *probabilistic unfolding analysis*, with principal component analysis and correspondence analysis. This produces techniques with results that can be interpreted both in probabilistic and in geometric terms. Moreover we propose algorithms that scale well, in the sense that they can be fitted efficiently to large matrices.

2. Problem

We fit an observed binary data matrix P to a predicted matrix $\Pi(X, Y)$. The predicted matrix, with elements in the open interval $(0, 1)$, is a function of X, an $n \times r$ matrix of *row scores*, and of Y, an $m \times r$ matrix of *column scores*. The parameter r is the *dimensionality* of the solution. The precise functional form of Π is specified below.

The computational problem we study in this paper is to minimize the distance between P and $\Pi(X, Y)$ over X and Y, where distance is measured by the loss function

$$\mathcal{D}(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{n}[p_{ij}\log\pi(x_i'y_j) + (1-p_{ij})\log(1-\pi(x_i'y_j))], \tag{1}$$

We discuss two different ways to specify the function π that maps the parameters in X and Y to the zero-one scale of the outcomes. In the *logit* case $\pi(x)$ is

$$\Psi(x) = \int_{-\infty}^{x}\psi(t)dt = \frac{1}{1+\exp\{-x\}}, \tag{2a}$$

where

$$\psi(x) = \frac{\exp\{-x\}}{(1+\exp\{-x\})^2} \tag{2b}$$

is the *standard logistic density function*. In the *probit* case $\pi(x)$ is

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt, \tag{3a}$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}t^2\}, \tag{3b}$$

the *standard normal density function*.

By defining a matrix $\Lambda = \{\lambda_{ij}\}$ with *logits* or *probits*, i.e. $\lambda_{ij} = \Psi^{-1}(\pi(x_i'y_j))$ or $\lambda_{ij} = \Phi^{-1}(\pi(x_i'y_j))$, we can write the basic relationship we fit as $\Lambda = XY'$. This shows that we are dealing with a fixed rank approximation problem on the logit or probit scale, a problem that is usually solved by *principal component analysis (PCA)* or, equivalently, *singular value decomposition (SVD)* in the linear case in which Λ is observed directly.

2.1. **Discussion.** The usual way to motivate loss function (1) is to assume that the z_{ij} are outcomes of independent Bernoulli trials with expected success $\pi(x_i'y_j)$. Then $\mathcal{D}$, except for irrelevant constants, is the *the negative log-likelihood* and minimizing $\mathcal{D}$ produces *maximum likelihood estimates*.

The second motivation, which seems more straightforward and natural in many actual data analysis situations, is that we want to find an approximate solution to the system of strict inequalities

$$x_i'y_j > 0 \qquad \forall p_{ij} = 1, \tag{4a}$$

$$x_i'y_j < 0 \qquad \forall p_{ij} = 0. \tag{4b}$$

It is easy to see that if the system (4) has a solution, then minimizing $\mathcal{D}$ will find it, and the minimum of $\mathcal{D}$ in that case will be zero. Conversely, we can only make $\mathcal{D}$ converge to zero by letting X and Y converge to a solution of (4). In fact what minimizing $\mathcal{D}$ is trying to achieve is

$$x_i'y_j \rightarrow \infty \qquad \forall p_{ij} = 1, \tag{5a}$$

$$x_i'y_j \rightarrow -\infty \qquad \forall p_{ij} = 0, \tag{5b}$$

although it will generally not succeed in its goal (only if the system (4) is solvable).

It is worth emphasizing that our loss function is, by definition, non-robust. The algorithm does not hesitate to move points to infinity if that makes loss smaller. Thus we tend to create outliers, certainly in small datasets. This is not necessarily a problem. Robustness is desirable if we are trying to estimate population characteristics, but not necessarily when we are trying to describe salient characteristics of a specific dataset. Thus our methods are very different from recent versions of robust PCA, such as Hubert et al. [2002]; Pison et al. [2003]. In our algorithm we identify X by setting $X'X = I$. This means that making π_{ij} as much like p_{ij} will tend to make Y large. Since we are mainly interested in the directions defined by the rows of Y this is not really a problem.

3. Algorithm

We develop a majorization algorithm, based on bounding the second derivative of the likelihood function. See Böhning and Lindsay [1988]; Böhning [1992]; Lange et al. [2000] for other examples in a logistic context. The general theory of majorization algorithms is reviewed briefly in Appendix A, and the basic majorization of the logit or probit log-likelihood is in Appendix B.

To treat both logit and probit cases simultaneously, we define

$$g_{ij}(x) = -\frac{\partial[p_{ij}\log\pi(x) + (1-p_{ij})\log(1-\pi(x)]}{\partial x}$$

and a matrix $G(X, Y)$ with elements $g_{ij}(x_i'y_j)$. From Appendix B we see that for the logit case

$$g_{ij}(x_i'y_j) = \Psi(x_i'y_j) - p_{ij},$$

while for the probit case

$$g_{ij}(x_i'y_j) = -\frac{p_{ij} - \Phi(x_i'y_j)}{\Phi(x_i'y_j)(1-\Phi(x_i'y_j))}\phi(x_i'y_j) =$$

$$= \begin{cases} -\frac{\phi(x_i'y_j)}{\Phi(x_i'y_j)} & \text{if } p_{ij} = 1, \\ \frac{\phi(x_i'y_j)}{1-\Phi(x_i'y_j)} & \text{if } p_{ij} = 0. \end{cases}$$

Moreover we write ω for the reciprocal of the upper bound on the second derivatives. For the logit case $\omega = 4$ and for the probit case $\omega = 1$.

Now for the algorithm. Suppose $X^{(k)}$ and $Y^{(k)}$ are the current best solution. We update them to find a better solution in two steps, similar to the E-step and the M-step in the EM-algorithm.

Algorithm 3.1 (Majorization)**.** *Start with some* $X^{(0)}$ *and* $Y^{(0)}$*.*

Step k(1): *Compute the matrix*

$$H^{(k)} = X^{(k)}\{Y^{(k)}\}' - \omega G(X^{(k)}, Y^{(k)})$$

Step k(2): *Solve the least squares matrix approximation problem*

$$\min_{X,Y} \mathbf{tr}\ (H^{(k)} - XY')'(H^{(k)} - XY')$$

by using the singular value decomposition (SVD).

Theorem 3.2. *The majorization algorithm 3.1 produces a decreasing sequence* $\mathcal{D}(X^{(k)}, Y^{(k)})$ *of loss function values, and all accumulation points of the sequence* $(X^{(k)}, Y^{(k)})$ *of iterates are stationary points.*

Proof. By the results in Appendix B

$$\mathcal{D}(X,Y) \leq \mathcal{D}(X^{(k)}, Y^{(k)}) + \frac{1}{2\omega}\mathbf{tr}\ (H^{(k)} - XY')'(H^{(k)} - XY') - \frac{\omega}{2}\mathbf{tr}\ G(X^{(k)}, Y^{(k)})'G(X^{(k)}, Y^{(k)}). \tag{6}$$

Only the middle term on the right hand side depends on X and Y and thus we if minimize this middle term to define $(X^{(k+1)}, Y^{(k+1)})$ we decrease loss. Now apply the general majorization results in Appendix A. □

4. Implementation Details

4.1. **Initial Estimate.** In the *R* implementation, given in Appendix D, the initial estimate for X and Y is simply taken as zero. This will obviously not be very good, and we may get some improvement by using homogeneity analysis De Leeuw [2003a]. The SVD majorization algorithm converges very fast in the initial steps, and then slows down to its slow linear rate, so these improvements will presumably be not very large.

Also observe that, both in the logit and in the probit case, starting with X and Y equal to zero, means that the first iteration computes the singular value decomposition of a matrix with element $p_{ij} - \frac{1}{2}$, and this will be already be close to the homogeneity analysis solution.

4.2. **Main Effects.** In some applications, for instance in random quadratic utility roll call models, the first column of X is restricted to consist of ones. More generally, we can fit

$$\lambda_{ij} = \mu + \alpha_i + \beta_j + \sum_{s=1}^{p} x_{is} y_{js}$$

which has both row and column main effects. In psychometrics this is sometimes called FANOVA [Gollob, 1968]. It has been studied in considerable detail by Gabriel and Gower in the context of biplot analysis [Gower and Hand, 1996].

Identification analysis of FANOVA suggests that if α is part of the specification, then we require that the columns of X sum to zero, and if β is part of the specification, then the columns of Y sum to zero. If μ is in the specification, then we also require that both α and β sum to zero.

In our algorithm, this amounts to centering the matrix $\tilde{Z}$ over rows and/or columns before computing the SVD. Clearly this does not really make the algorithm any more complicated.

It is perhaps worth saying here that the specification with only main effects and no interaction terms is the Rasch model [Fischer and Molenaar, 1995]. Moreover, we can easily implement the constrained forms of PCA discussed by Takane and his co-workers [Takane and Shibayama, 1991; Takane et al., 1995; Takane and Hunter, 2001].

4.3. **Inner Iterations for Missing Data.** If there are missing data then the matrix approximation problem becomes

$$\min_{X,Y} \sum \{(h_{ij}^{(k)} - x_i' y_j)^2 \mid (i, j) \in N\},$$

where N is the subset of non-missing index pairs.

We now use the familiar least squares augmentation trick, used in non-balanced ANOVA by Yates and Wilkinson and in least squares factor analysis by Thomson and Harman. See De Leeuw [1994]; De Leeuw and Michailidis [1999] for references and for further discussion of augmentation.

We define inner iterations in each iteration of our majorization algorithm to impute the missing data. The inner iterations start with $x_i^{(k,0)} = x_i^{(k)}$ and $y_j^{(k,0)} = y_j^{(k)}$.

$$\tilde{h}^{(k,\ell)} = \begin{cases} h_{ij}^{(k)} & \text{if } (i,j) \in N, \\ \{x_i^{(k,\ell)}\}' y_j^{(k,\ell)} & \text{if } (i,j) \notin N, \end{cases}$$

We then do an SVD to find $X^{(k,\ell+1)}$ and $Y^{(k,\ell+1)}$, and continue the inner iterations. Actually, in our R implementation in Appendix D we only perform a single inner iteration, which basically means that we always perform a singular value decomposition on $\tilde{H}^{(k,0)}$ which is just our previous $H^{(k)}$ with missing elements imputed by setting them to the corresponding elements of $\{X^{(k)}\}'Y^{(k)}$.

4.4. **Innermost Iterations for the SVD.** It may not be a good idea to do a complete SVD after computing a new H or $\tilde{H}$, even if we use an SVD algorithm that only computes p singular vectors. We could use an iterative SVD method such as the simultaneous iteration method first proposed by Daugavet [1968], and only perform one or a small number of innermost iterations before updating $\tilde{H}$. This may ultimately lead to fewer computations. But observe that going this way is probably mainly relevant if the algorithm is written in a compiled language such as C, writing our own innermost iterations in a interpreted language such as R with fast compiled SVD operators will most likely slow down the process.

Each Daugavet iteration

$$X \leftarrow \tilde{Z}Y(Y'Y)^{-1},$$
$$Y \leftarrow \tilde{Z}'X(X'X)^{-1},$$

basically requires two matrix multiplications, so even for big matrices it is quite inexpensive. To identify along the way, the iterations are typically implemented as

$$X \leftarrow \mathbf{orth}(ZY),$$
$$Y \leftarrow \tilde{Z}'X,$$

where **orth** is an orthogonalization method such as Gram-Schmidt or QR. This makes the method identical to the Bauer-Rütishauser simultaneous iteration method, used in a similar context by Gifi [1990, page 98-99].

If an iterative SVD method is implemented, then we have to distinguish the outer iterations of the majorization algorithm, the inner iterations of the augmentation method to impute missing values, and the innermost iterations to compute or improve the SVD. The number of inner and innermost iterations will influence the amount of computation in an outer iteration and the convergence speed of the algortithm.

4.5. **Factor Analysis.** It is easy to adapt our algorithm to fitting factor analysis instead of principal component analysis decompositions. We use the basic setup of De Leeuw [2003b]. Thus, instead of fitting

$$\underset{n\times m}{\Lambda} \approx \underset{n\times p}{X}\;\underset{p\times m}{Y'}$$

where $X'X = I$, we fit

$$\underset{n\times m}{\Lambda} \approx \underset{n\times p}{X}\;\underset{p\times m}{Y'} + \underset{n\times m}{E}\;\underset{m\times m}{D}.$$

where $X'X = I$, $X'E = 0$, $E'E = I$ and D is diagonal. In the roll-call context, this means we distinguish a common space of roll calls and in addition a unique dimension for each roll call. In constructing combines the majorization method proposed here with the alternating least squares inner iterations of De Leeuw [2003b] that replace the SVD.

5. Roll calls

There have been very interesting recent developments in multidimensional roll call analysis. Let us first outline the basic way of thinking in the field [Clinton et al., 2003]. We work in $\mathbb{R}^r$. Each legislator has an ideal point x_i in this space and each roll call has both a yes-point u_j and a no-point v_j. The utlities for legislator i to vote "yes" or "no" on roll call j have both a fixed and a random component. We use the Dutch Convention [Hemelrijk, 1966] to underline random variables. Thus

$$\begin{aligned}\underline{\xi}_{ij}^1 &= \Gamma(x_i, u_j) + \underline{\epsilon}_{ij}^1,\\ \underline{\xi}_{ij}^0 &= \Gamma(x_i, v_j) + \underline{\epsilon}_{ij}^0,\end{aligned}$$

where Γ is some utility function defined on pairs of points. This means that the legislator will vote "yes" if $\underline{\xi}_{ij}^1 > \underline{\xi}_{ij}^0$, i.e. when

$$-(\underline{\epsilon}_{ij}^1 - \underline{\epsilon}_{ij}^0) < \Gamma(x_i, u_j) - \Gamma(x_i, v_j)$$

If F is the cumumative probability distribution of $-(\underline{\epsilon}_{ij}^1 - \underline{\epsilon}_{ij}^0)$, then π_{ij}, the probability that legislator i will vote "yes" on roll call j is

$$\pi_{ij} = F\{\Gamma(x_i, u_j) - \Gamma(x_i, v_j)\}.$$

Clearly we still have a lot of choices to make in this general setup, because we can specify both Γ and F in many ways.

Suppose, for instance, we use the *bilinear utilities*, with $\Gamma(x, y) = x'y$. Then

$$\Gamma(x_i, u_j) - \Gamma(x_i, v_j) = x_i'(u_j - v_j).$$

If we use *quadratic utilities* [Poole, 2001], then $\Gamma(x, y) = -\|x - y\|^2$ and thus

$$\Gamma(x_i, u_j) - \Gamma(x_i, v_j) = 2x_i'(u_j - v_j) - (\|u_j\|^2 - \|v_j\|^2).$$

Clearly the bilinear and quadratic specification cannot be distinguished [Böckenholt, in press] and both can be written as $\pi_{ij} = F\{x_i'y_j\}$.

In what has been, at least until recently, the most popular and most sophisticated approach to multidimensional roll call modeling Poole and Rosenthal [1985, 1997] assume that

$$\Gamma(x, y) = \zeta \exp\{-0.125\|x - y\|^2\}$$

and that F is the logistic cdf. Poole and Rosenthal argue for the advantages of this *Gaussian utility*, but clearly it is more complicated than the quadratic form. The Gaussian utility and the logistic distribution define the NOMINATE model, and the parameters are fitted by a complicated but seemingly effective block relaxation [De Leeuw, 1994] optimization of the likelihood function.

Of course quadratic utility goes back to at least unfolding theory, invented by Coombs in the fifties and summarized in his book [Coombs, 1964]. In roll call analysis Poole [1999] has gone back recently to the Coombsian roots of the quadratic utility model. He has designed a more geometrical and more heuristic procedure to minimize the number of misclassifications resulting from fitting the quadratic utility model.

Also recently, the basic quadratic roll call model has been cast in a Bayesian framework and fitted with Markov Chain Monte Carlo (MCMC) methods [Jackman, 2001]. Since typically flat priors are used, estimates will tend to be similar to the maximum likelihood estimates. We have no comparisons of the relative speed or behavior of majorization and MCMC algorithms, but the general considerations can at least suggest some differences. Convergence of both procedures will be slow, convergence of the majorization algorithm will be more regular and smooth, and the Bayesian computations will more easily give information about stability. The Bayesian framework clearly aims to produce more more than just a nice picture for data reduction [Jackman, 2000] and, if one believes the assumptions on which the Bayesian computations and interpretations are build, indeed it does.

The approach taken by De Leeuw [2003a] is quite the opposite of the Bayes/MCMC tandem. Various measures of the size of a cloud of points in $\mathbb{R}^r$ are considered, and a picture is constructed in such a way that the average of the sizes of the yes-clouds and the no-clouds over issues is minimized. If cloud size is defined as squared distance to the centroid of the cloud this leads to multiple correspondence analysis [Greenacre, 1984; Gifi, 1990], a technique which is computationally a relatively simple technique, because it requires computation of just one single SVD.

Clearly the technique in this paper combines aspects of the quadratic utility approach and the singular value approach. Our emphasis is on data reduction, not on inference, although it is possible to use standard techniques to compute confidence regions. All the needed derivatives of the likelihood function have been computed by Rivers [2003].

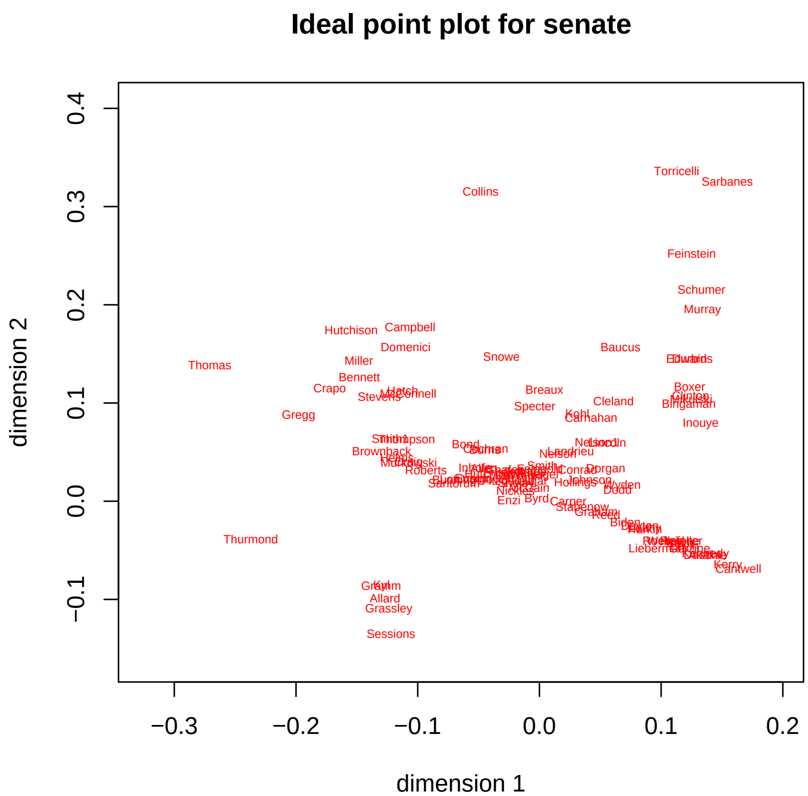
6. Examples

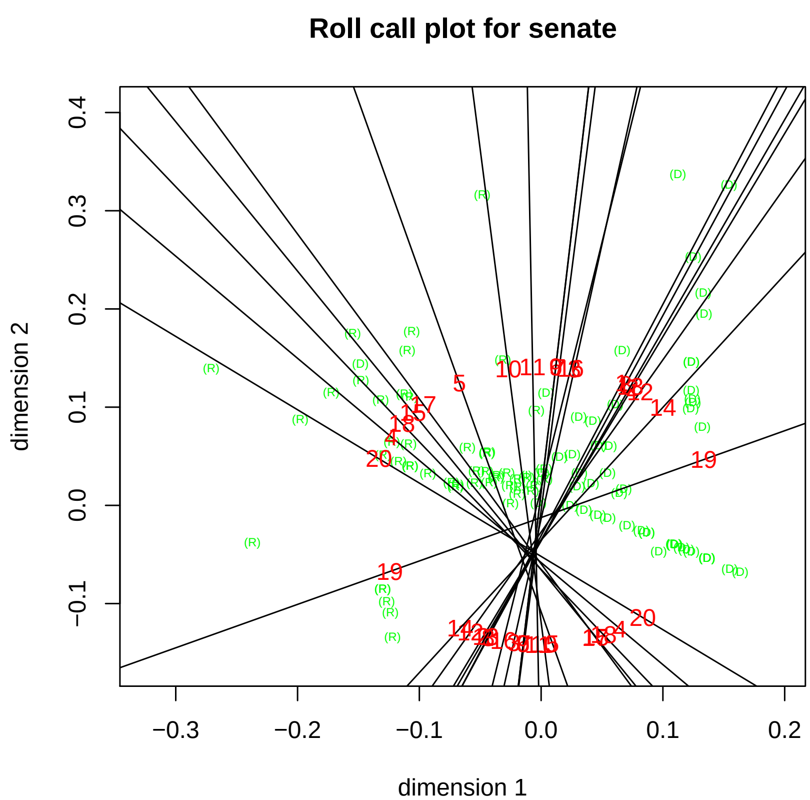
6.1. **Senate.** We analyze 2001 senate votes on 20 issues selected by Americans for Democratic Action [Ada, 2002]. Descriptions of the roll calls are given in Appendix C. We use the logit function.

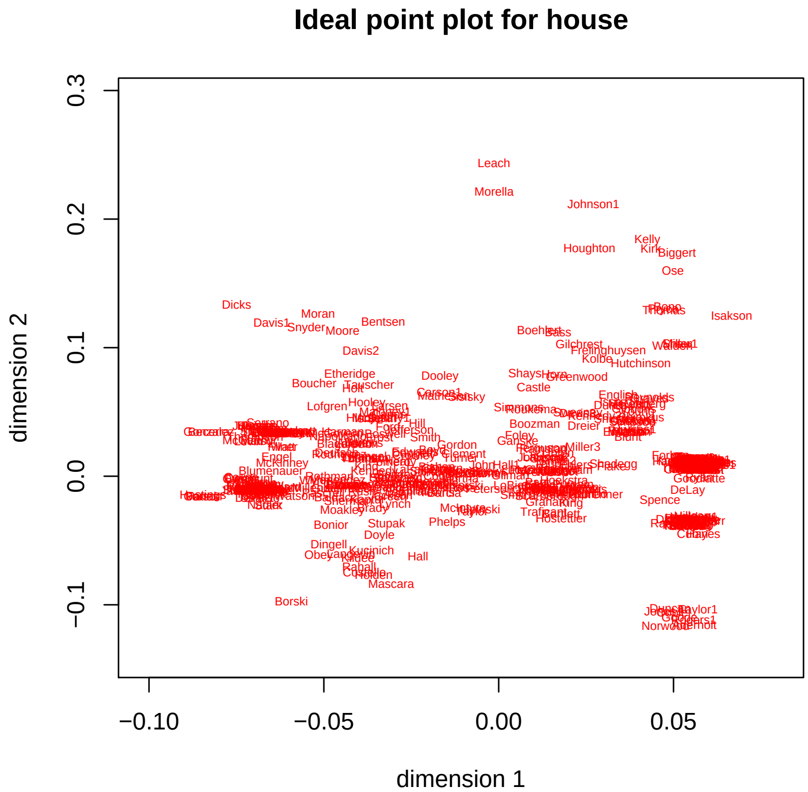
We start with setting all parameters equal to zero. The algorithm then takes 997 iterations to stabilize the negative log-likelihood to three decimal places. At the solution the proportion of correctly classified votes is 0.9425. After one iteration it is 0.9228, after 100 iterations it is 0.9384. The analysis illustrates that the ML method tries to make as many fitted probabilities π equal to zero and one (i.e. equal to the corresponding p). In this example 75% is very close to either zero or one.

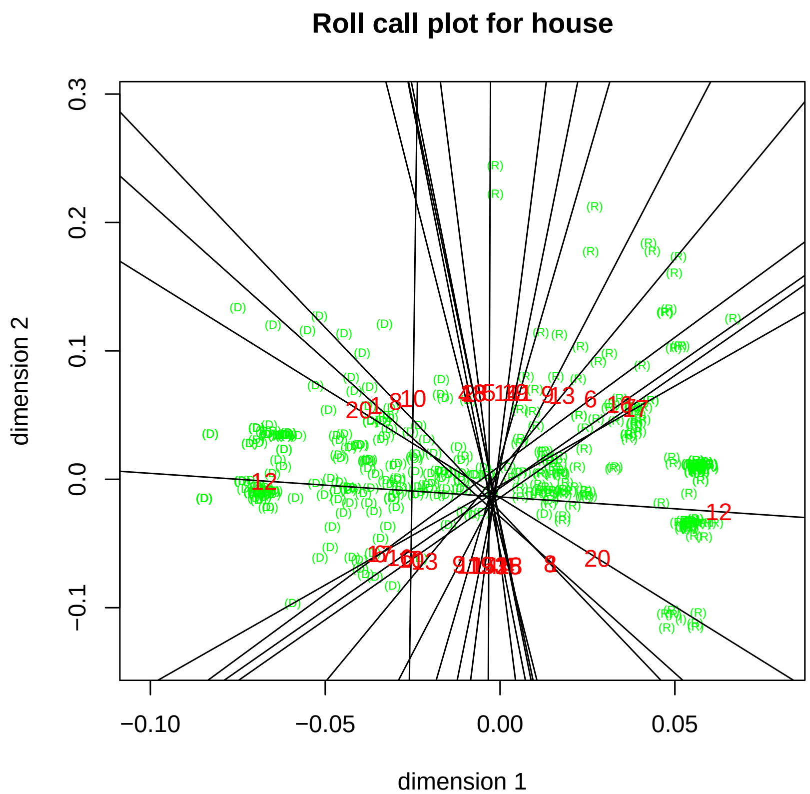
6.2. **House.** We also use the data for the 2001 House from Ada [2002], with twenty different role calls. Now the algorithm needs 186 iterations to attain three decimals precision. Again the logit function is fitted. The proportion of correct classifications is 0.9387, up just a tiny bit from 0.9322 after the first iteration. The number of fitted probabilities which are indistinguishable from zero or one is 70%.

6.3. **Results.** Results are in the figure below, both for the Senate and the House. The legislators in the legislator plots are labeled by name. The roll call plots have both the legislators (now labeled by party) as points and the roll calls as directions. Each roll call separates the legislators into two halfspaces, containing the "yes" and "no" groups, and we can easily count the misclassification errors.


Ideal point plot for senate
dimension 1
dimension 2
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
Torricelli
Sarbanes
Collins
Feinstein
Schumer
Murray
Hutchison
Campbell
Domenici
Snowe
Baucus
Thomas
Miller
Bennett
Crapo
Breaux
Cleland
Specter
Gregg
Inouye
Thurmond
Allard
Grassley
Sessions
Cantwell


Roll call plot for senate
dimension 1
dimension 2
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4


Ideal point plot for house
dimension 2
dimension 1
−0.10
−0.05
0.00
0.05
−0.1
0.0
0.1
0.2
0.3


Roll call plot for house
dimension 2
dimension 1
−0.10
−0.05
0.00
0.05
−0.1
0.0
0.1
0.2
0.3

APPENDIX A. MAJORIZATION METHODS

A.1. **General Principles.** The algorithms proposed in this paper are all of the majorization type. Majorization is discussed in general terms in De Leeuw [1994]; Heiser [1995]; Lange et al. [2000].

In a majorization algorithm the goal is to minimize a function $\phi(\theta)$ over $\theta \in \Theta$, with $\Theta \subseteq \mathbb{R}^p$. Suppose that a function $\psi(\theta, \xi)$ defined on $\Theta \times \Theta$ satisfies

$$\phi(\theta) \leq \psi(\theta, \xi) \text{ for all } \theta, \xi \in \Theta, \tag{7a}$$

$$\phi(\theta) = \psi(\theta, \theta) \text{ for all } \theta \in \Theta. \tag{7b}$$

Thus, for a fixed ξ, $\psi(\bullet, \xi)$ is above ϕ, and it touches ϕ at the point $(\xi, \phi(\xi))$. We then say that $\psi(\theta, \xi)$ *majorizes* $\phi(\theta)$ at ξ.

There are two key theorems associated with these definitions.

Theorem A.1. *If ϕ attains its minimum on Θ at $\hat{\theta}$, then $\psi(\bullet, \hat{\theta})$ also attains its minimum on Θ at $\hat{\theta}$.*

Proof. Suppose $\psi(\tilde{\theta}, \hat{\theta}) < \psi(\hat{\theta}, \hat{\theta})$ for some $\tilde{\theta} \in \Theta$. Then, by (7a) and (7b), $\phi(\tilde{\theta}) \leq \psi(\tilde{\theta}, \hat{\theta}) < \psi(\hat{\theta}, \hat{\theta}) = \phi(\hat{\theta})$, which contradicts the definition of $\hat{\theta}$ as the minimizer of ϕ on Θ. □

Theorem A.2. *If $\tilde{\theta} \in \Theta$ and $\hat{\theta}$ minimizes $\psi(\bullet, \tilde{\theta})$ over Θ, then $\phi(\hat{\theta}) \leq \phi(\tilde{\theta})$.*

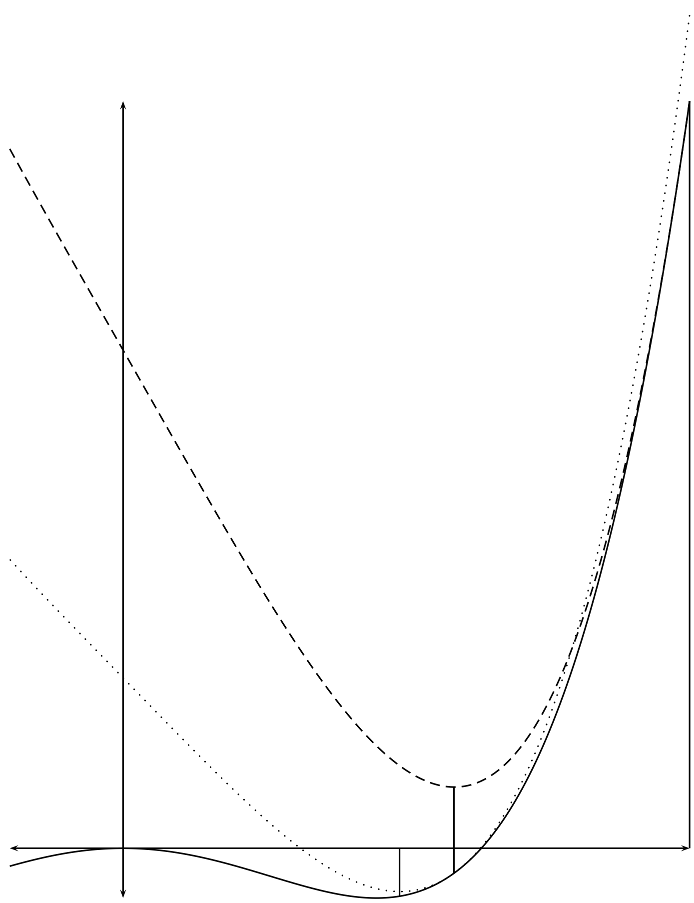
Proof. By (7a) we have $\phi(\hat{\theta}) \leq \psi(\hat{\theta}, \tilde{\theta})$. By the definition of $\hat{\theta}$ we have $\psi(\hat{\theta}, \tilde{\theta}) \leq \psi(\tilde{\theta}, \tilde{\theta})$. And by (7b) we have $\psi(\tilde{\theta}, \tilde{\theta}) = \phi(\tilde{\theta})$. Combining these three results we get the result. □

These two results suggest the following iterative algorithm for minimizing $\phi(\theta)$. Suppose we are at step k.

Step 1:: Given a value $\theta^{(k)}$ construct a majorizing function $\psi(\theta^{(k)}, \xi)$.
Step 2:: Minimize $\psi(\theta^{(k)}, \xi)$ with respect to ξ. Set $\theta^{(k+1)} = \xi^{\max}$.
Step 3:: If $|\phi(\theta^{(k+1)} - \phi(\theta^{(k)}| < \epsilon$ for some predetermined $\epsilon > 0$ stop; else go to Step 1.

In order for this algorithm to be of practical use, the majorizing function ψ needs to be easy to minimize, otherwise nothing substantial is gained by following this route. Notice, that in case we are interested to maximize ϕ, we have to find a minorizing function ψ that needs to be maximized in Step 2.

We demonstrate next how the idea behind majorization works with a simple example.



Example A.1. This is an artificial example, chosen for its simplicity. Consider $\phi(\theta) = \theta^4 - 10\theta^2,\ \theta \in \mathbb{R}$. Because $\theta^2 \geq \xi^2 + 2\xi(\theta - \xi) = 2\xi\theta - \xi^2$ we see that $\psi(\theta, \xi) = \theta^4 - 20\xi\theta + 10\xi^2$ is a suitable majorization function. The majorization algorithm is $\theta^+ = \sqrt[3]{5\xi}$.

The algorithm is illustrated in Figure A.1. We start with $\theta(0) = 5$. Then $\psi(\theta, 5)$ is the dashed function. It is minimized at $\theta^{(1)} \approx 2.924$, where $\psi(\theta^{(1)}, 5) \approx 30.70$, and $\phi(\theta^{(1)}) \approx -12.56$. We then majorize by using the dotted function $\psi(\theta, \theta^{(1)})$, which has its minimum at about 2.44, equal to about -21.79. The corresponding value of ϕ at this point is about -24.1. Thus we are rapidly getting close to the local minimum at $\sqrt{5}$, with value 25. The linear convergence rate at this point is $\frac{1}{3}$.

We briefly address next some convergence issues (for a general discussion see the book by Zangwill [1969] and also Meyer [1976]). If ϕ is bounded above (below) on Θ, then the algorithm generates a bounded increasing sequence of function values $\phi(\theta^{(k)}$, thus it converges to $\phi(\theta^{\infty})$. For example, continuity of ϕ and compactness of Θ would suffice for establishing the result. Moreover with some addiitonal mild continuity considerations [De Leeuw, 1994] we get that $||\theta^{(k)}-\theta^{(k+1)}|| \rightarrow 0$, which in turn implies, because of a result by Ostrowski [1966], that either θ converges to a stable point or that there is a continuum of limit points (all with the same function value). Hence, majorization algorithms for all practical purposes find local optima.

We make two final points about this class of algorithms. It is not necessary to actually minimize the majorization function in each step, it suffices to decrease it in a systematic way, for instance by taking a single step of a convergent "inner" iterative algorithm. And the rate of convergence of majorization algorithms is generally linear, in fact it is equal to the size of the second derivatives of the majorization function compared to the size of the second derivatives of the original function [De Leeuw and Michailidis, 1999].

Appendix B. Quadratic Majorization of Negative Log Likelihood

B.1. **The Logit Case.** Define

$$f(x) = -p\log\pi(x) - (1-p)\log(1-\pi(x)),$$

where

$$\pi(x) = \frac{1}{1+\exp(-x)}$$

Theorem B.1. *f is strictly convex on $(0,1)$ and has a uniformly bounded second derivative satisfying $0 < f''(x) < \frac{1}{4}$.*

Proof. Simple calculation gives

$$f'(x) = \pi(x) - p,$$
$$f''(x) = \pi(x)(1-\pi(x)).$$

Clearly

$$0 < f''(x) < \frac{1}{4}$$

for all $0 < x < 1$, which is all we need. □

Theorem B.2. *Let*

$$g(x, y) = f(y) + \frac{1}{8}[x - (y - 4(\pi(y) - p))]^2 - 2(\pi(y) - p)^2$$

Then g majorizes f in the sense that

$$f(x) \leq g(x, y) \quad \forall x, y,$$
$$f(x) = g(x, x) \quad \forall x.$$

Proof. From Theorem B.1 we know

$$f(x) \leq f(y) + (\pi(y) - p)(x - y) + \frac{1}{8}(x - y)^2$$

By completing the square we see that the right hand side is $g(x, y)$. □

B.2. **The Probit case.** Define

$$f(x) = -p \log \Phi(x) - (1 - p) \log(1 - \Phi(x)),$$

where

$$\Phi(x) = \int_{-\infty}^{x} \phi(z) dz,$$

and

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}z^2\}.$$

Theorem B.3. *The function f is strictly convex on* $(0, 1)$ *and has a uniformly bounded second derivative satisfying* $0 < f''(x) < 1$.

Proof. By simple computation

$$f'(x) = -\frac{p - \Phi(x)}{\Phi(x)(1 - \Phi(x))}\phi(x)$$

and

$$f''(x) = \frac{p - \Phi(x)}{\Phi(x)(1 - \Phi(x))}x\phi(x) + \phi^2(x)\frac{p + \Phi^2(x) - 2p\Phi(x)}{\Phi^2(x)(1 - \Phi(x))^2}$$

If $p = 1$ we have $f'(x) = -m(x)$, where

$$m(x) = \frac{\phi(x)}{\Phi(x)},$$

is the Inverse Mills' Ratio, and

$$f''(x) = xm(x) + m^2(x).$$

We now use a trick from Sampford [1953]. Consider a standard normal random variable, truncated on the right (from above) at x. Its variance is $1-xm(x)-m^2(x)$, see Johnson et al. [1994, section 10.1], and because variance is positive, we see that $f''(x) < 1$. On the other hand, the variance must be less than that of the standard normal, which implies $f''(x) > 0$.

If $p = 0$ then $f'(x) = M(x)$, where

$$M(x) = \frac{\phi(x)}{1 - \Phi(x)},$$

is Mills' Ratio, and

$$f''(x) = -xM(x) + M^2(x).$$

Consider a standard normal random variable, truncated on the left (from below) at x. Its variance is $1 + xM(x) - M^2(x)$. Again this implies $0 < f''(x) < 1$.

The second derivative $f''(x)$ is linear in p for fixed x. Since it is less than one and larger than zero for both $p = 0$ and $p = 1$, it must also be less than one and larger than zero for all intermediate values of p. □

Theorem B.4. *Let*

$$g(x, y) = f(y) + \frac{1}{2}\left[(x - (y + \frac{p - \Phi(y)}{\Phi(y)(1 - \Phi(y))}\phi(y))\right]^2 - \frac{1}{2}\left[\frac{p - \Phi(y)}{\Phi(y)(1 - \Phi(y))}\phi(y)\right]^2$$

Then g majorizes f in the sense that

$$f(x) \leq g(x, y) \quad \forall x, y,$$
$$f(x) = g(x, x) \quad \forall x.$$

Proof. From Theorem B.3 we know

$$f(x) \leq f(y) - \frac{p - \Phi(y)}{\Phi(y)(1 - \Phi(y))}\phi(y)(x - y) + \frac{1}{2}(x - y)^2$$

By completing the square we see that the right hand side is $g(x, y)$. □

APPENDIX C. ADA ROLL CALL DESCRIPTIONS

C.1. **Senate.** The votes selected cover a full spectrum of domestic, foreign, economic, military, environmental and social issues. We tried to select votes which display sharp liberal/conservative contrasts. In many instances we have chosen procedural votes: amendments, motions to table, or votes on rules for debate. Often these votes reveal true attitudes frequently obscured in the final votes.

(1) **Ashcroft Attorney General Confirmation.** Confirmation of President Bush's nomination of John Ashcroft of Missouri to serve as U.S. Attorney General. Confirmed 58-42. Feb. 1, 2001. A no vote is a +.
(2) **SJ Res 6. Ergonomics Rule Disapproval.** Passage of a joint resolution to reverse the ergonomics workplace safety rule submitted by the Clinton Administration's Labor Department. Passed 56-44. March 6, 2001. A no vote is a +.
(3) **S 420. Social Security "Lockbox".** Domenici (R-NM) motion to waive the Budget Act in order to ensure that the Social Security surplus is used only to pay down the public debt until Social Security reform legislation is enacted. The bill would also ensure that the surplus in the Medicare Hospital Insurance Trust Fund is used only to pay down the public debt until Medicare reform legislation is enacted. Motion rejected 52-48 (a three-fifths majority vote - 60 - is required to waive the Budget Act.) March 13, 2001. A no vote is a +.
(4) **S 27. Campaign Finance Reform.** McCain (R-AZ) motion to kill the Hatch (R-UT) amendment requiring unions and corporations to obtain permission from individual dues-paying workers or shareholders before spending money on political activities. The Hatch amendment was intended as a "poison pill" that, if passed and attached to the campaign finance reform bill, would destroy any chances the full reform bill had of passage. The Hatch amendment would also require corporations and unions to disclose information regarding the funds spent on political activities. Motion agreed to 69-31. March 21, 2001. A yes vote is a +.
(5) **S 27. Soft Money Cap.** McCain (R-AZ) motion to kill a Hagel (R-NE) amendment to limit at $60,000 per year soft money contributions by individuals, political action committees, corporations and unions to national and state political party committees. The Hagel amendment would render

the underlying reform bill's ban on soft money ineffective. Motion agreed to 60-40. March 27. 2001. A yes vote is a +.

(6) **H Con Res 83. Prescription Drug Benefit/Tax Cuts.** Grassley (R-IA) amendment to reserve $300 billion over 10 years to create a Medicare prescription drug benefit and overhaul the program. This amendment was a response to Democratic legislation which would have allocated $311 billion for the benefit and not allowed the benefit's funding to come from the Medicare Hospital Trust Fund Surplus. Adopted 51-50, with Vice President Cheney casting a "yea" vote. April 3, 2001. A no vote is a +.

(7) **H Con Res 83. Fiscal 2002 Budget Reconciliation.** Domenici (R-NM) amendment to instruct the Senate Finance Committee to report two reconciliation bills to the Senate that would reduce revenue levels by not more than the President's proposed $1.6 trillion tax cut, and include a $60 billion economic stimulus package for fiscal 2001. Adopted 51-49. April 5, 2001. A no vote is a +.

(8) **H Con Res 83. Funding for Environmental Programs.** Corzine (D-NJ) amendment to increase funding for a wide variety of environmental programs by $50 billion and set aside $50 billion for debt reduction. The increases would be offset by reductions in the proposed tax cut. Rejected 46-54. April 5, 2001. A yes vote is a +.

(9) **H Con Res 83. "Marriage Penalty" Tax.** Hutchison (R-TX) amendment to increase the proposed tax cut by $69 billion for fiscal 2002-2011 in an effort to eliminate the co-called marriage penalty. Adopted 51-50, with Vice President Cheney casting a "yea" vote. April 5, 2001. A no vote is a +.

(10) **H Con Res 83.** Disabilities Education Act Funding. Breaux (D-LA) amendment to redirect $70 billion from the proposed tax cut to funding for the Individuals with Disabilities Education Act (IDEA) over 10 years. Adopted 54-46. April 5, 2001. A yes vote is a +.

(11) **S I. School Renovation and Construction.** Harkin (D-IA) amendment to authorize $1.6 billion for fiscal 2002 and such sums as necessary for each fiscal year between 2003 and 2006 for the construction and renovation of public elementary and secondary school buildings. Rejected 49-50. May 16, 2001. A yes vote is a +.

(12) **HR 1836. Estate Tax.** Dorgan (D-ND) amendment to strike the estate-tax repeal provision and repeal the estate tax in 2003 for only all qualified

family-owned farms and businesses. It also would reduce the top estate-tax rate bracket to 45 percent. Rejected 43-56. May 21, 2001. A yes vote is a +.

(13) **HR 1836. Head Start.** Kennedy (D-MA) amendment to condition the reductions in the marginal income-tax rate on full funding for Head Start programs. Motion rejected 45-54. May 22, 2001. A yes vote is a +.

(14) **HR 1836. Tax Cut Reconciliation Bill.** Adoption of the conference report on the bill to reduce taxes by $1.35 trillion through fiscal 2011 through income tax rate cuts, relief of the "marriage penalty," phase-out of the federal estate tax, doubling of the child tax credit, and new incentives for retirement savings. A new 10 percent tax rate would be created retroactive to January 1. The bill would double the $5000-per-child tax credit by 2010 and make it refundable, raise the estate tax exemption to $1 million in 2002 and repeal the tax in 2010, increase the standard deduction for married couples to double that of singles over five years, beginning in 2005, and increase annual contributions limits for Individual Retirement Accounts. The bill's provisions would expire December 31, 2010. Adopted 58-33. May 26, 2001. A no vote is a +.

(15) **S I. School Vouchers.** Gregg (R-NH) amendment to create a demonstration program in 10 school districts to provide public school children with federal funds (vouchers) to transfer to another public school or a private school, including religious schools. The amendment would authorize $50 million for fiscal 2002 and subsequent necessary sums for the next six fiscal years. Rejected 41-58. June 12, 2001. A no vote is a +.

(16) **S I. Boy Scouts/Anti-Discrimination.** Helms (R-NC) amendment to withhold federal education funds from public elementary and secondary schools that bar the Boy Scouts of America from using school facilities. The targeted schools bar the Boy Scouts because the organization discriminates against gay men. Adopted 51-49. June 14, 2001. A no vote is a +.

(17) **S 1052. Patients' Bill of Rights.** Passage of the bill to provide federal patient protections and allow patients to appeal a health maintenance organization's (HMO) decision on coverage and treatment. It also would allow patients to sue health insurers in state courts over quality-of-care claims and, at the federal level, over administrative or non-medical coverage disputes. Passed 59-36. June 29, 2001. A yes vote is a +.

(18) **HR 2299. NAFTA/Mexican Trucks.** Shelby (R-AL) motion to uphold a border truck inspection program which allows Mexican trucks to receive three-month permits if they pass safety inspections. The motion also upholds a grant of $60 billion to the Transportation Department and various agencies. Motion agreed to 65-30. July 27, 2001. A yes vote is a +.
(19) **S 1438. Military Base Closures.** Warner (R-VA) motion to authorize an additional round of U.S. military base realignment and closures in 2003. Motion agreed to 53-47. September 25, 2001. A yes vote is a +.
(20) **HR 2944. Fiscal 2002 District of Columbia Appropriations.** Passage of the bill to provide $408 million for the District of Columbia in fiscal 2002, including funds for the city's courts and corrections system and $16.1 million for an emergency response plan following the September 11 attacks. The bill also would approve a $7.2 billion budget for the District. Passed 75-24. November 7, 2001. A yes is a +.

C.2. **House.** The votes selected cover a full spectrum of domestic, foreign, economic, military, environmental and social issues. We tried to select votes which display sharp liberal/conservative contrasts. In many instances we have chosen procedural votes: amendments, motions to table, or votes on rules for debate. Often these votes reveal true attitudes frequently obscured in the final votes.

(1) **HR 333. Bankruptcy Overhaul.** Jackson-Lee (D-TX) amendment to allow debtors to deduct additional medical and child-care expenses before determining their eligibility for Chapter 7 bankruptcy status. The amendment also expands the definition of family farmer, changes the standards for calculating median income, and includes debtor privacy provisions. Rejected 160-258. March 1, 2001. A yes vote is a +.
(2) **SJ Res 6. Ergonomics Rule Disapproval.** Passage of the joint resolution to reverse the ergonomics workplace safety rule submitted by the Clinton Administration's Labor Department. Passed 223-206. March 7, 2001. A no vote is a +.
(3) **HR 3. Income Tax Reduction.** Passage of the White House's bill to lower federal income taxes by restructuring the five existing tax brackets into four - 10 percent, 15 percent, 25 percent and 33 percent. The benefits of this tax cut go disproportionately to the wealthy and to major corporations. The large cost of the legislation would jeopardize domestic spending programs

aimed at middle- and low-income Americans. Passed 230-198. March 8, 2001. A no vote is a +.

(4) **HR 6. Marriage Tax Reduction.** Rangel (D-NY) substitute amendment to reduce taxes by $585.5 billion through 2011. This tax cut would be considerably less regressive and more equitable than the Republican version. The Rangel plan would create a new 12 percent bracket for the first $20,000 of a couple's taxable income and $10,000 for single taxpayers. It also would increase the standard deduction for married couples filing jointly to twice that of individuals filing singly. Additionally, the amendment would simplify and expand the earned-income tax credit for low-income earners. Rejected 196-231. March 29, 2001. A yes vote is a +.

(5) **HR 8. Estate Tax Relief.** Rangel (D-NY) substitute amendment to increase the estate tax exemption from $675,000 to $2 million ($4 million for married couples) in 2002, rising to $2.5 million by 2010. This legislation serves as an alternative to the drastic Republican abolition of the progressive estate tax. The Rangel tax cut would lower federal revenue by $39.2 billion over ten years. The amendment would retain current-law ”step-up basis” provisions, and replace the credit for estate taxes paid to a state with a deduction. Rejected 201-227. April 4, 2001. A yes vote is a +.

(6) **HR 503. Fetal Protection.** Passage of the bill to make it a criminal offense to injure or kill a fetus during the commission of a violent federal crime. The measure would establish criminal penalties equal to those that would apply if the injury or death occurred to a pregnant woman, regardless of the perpetrator's knowledge of the pregnancy or intent to harm the fetus. The bill states that its provisions should not be interpreted to apply to consensual abortion or to a woman's actions with respect to her pregnancy. The death penalty could not be imposed under this bill. Passed 252-172. April 26, 2001. A no vote is a +.

(7) **HR 1. School Vouchers.** Armey (R-TX) amendment to provide federal funding for students to attend private schools, including religious schools, if they are currently enrolled in schools that are dangerous or have been low-performing for three years. Crime victims also would be provided with funding to attend alternative private schools. Rejected 155-273. May 23, 2001. A no vote is a +.

(8) **HR 1836. Tax Cut Reconciliation Bill.** Adoption of the conference report on the bill to reduce taxes by $1.35 trillion through fiscal 2011 via income

tax rate cuts, relief of the ”marriage penalty,” phaseout of the federal estate tax, doubling of the child tax credit, and new incentives for retirement savings. A new 10 percent tax rate would be created retroactive to January 1. The bill would: double the $500-per-child tax credit by 2010 and make it refundable; raise the estate tax exemption to $1 million in 2002 and repeal the tax in 2010; increase the standard deduction for married couples to double that of singles over five years, beginning in 2005; and increase annual contributions limits for Individual Retirement Accounts. The bill's provisions would expire December 31, 2010. Adopted 240-154. May 26, 2001. A no vote is a +.

(9) **HR 2356. Campaign Finance Reform.** Adoption of the rule to allow the House to consider a ban on ”soft money” donations to national political parties. This rule was crafted by campaign finance reform foes to disallow amendments which fine-tune the bill and, thus, keep reform advocates from gathering more votes in support of final passage. Beyond banning soft money, the original reform legislation would allow up to $10,000 in soft-money donations to state and local parties for voter registration and get-out-the vote activity. The reform bill would prevent issue ads from targeting specific candidates within 60 days of a general election or 30 days of a primary. Additionally, the legislation would maintain the current individual contribution limit of $1,000 per election for House candidates but raise it to $2,000 for Senate candidates, both of which would be indexed for inflation. Rejected 203-228. July 12, 2001. A no vote is a +.

(10) **HJ Res 36. Flag Desecration.** Passage of the joint resolution proposing a Constitutional amendment to prohibit physical desecration of the U. S. flag. Passed 298-125. (A two-thirds majority vote of those present and voting - 282 in this case - is required to pass a joint resolution proposing an amendment to the Constitution.) July 17, 2001. A no vote is a +.

(11) **HR 7. Faith-Based Initiative.** Conyers (D-MI) motion to recommit the bill to the Judiciary Committee with instructions to add language stating that federally-funded religious service providers cannot discriminate based on religion and that no provision supercedes state or local civil rights laws. Motion rejected 195-234. July 19, 2001. A yes vote is a +.

(12) **HJ Res 50. China Normalized Trade Relations.** Passage of a joint resolution to deny the President's request to provide normal trade relations (formerly known as most-favored-nation trade status) for items produced

in China from July 2001 through July 2002. Rejected 169-259. July 19, 2001. A yes vote is a +.

(13) **HR 4. ANWR Drilling Ban.** Markey (D-MA) amendment to maintain the current prohibition on oil drilling in the Arctic National Wildlife Refuge. Rejected 206-223. August 1, 2001. A yes vote is a +.

(14) **HR 2563. Patients' Rights/HMO Liability.** Norwood (R-GA) amendment to limit liability and damage awards when a patient is harmed by denial of health care. This amendment was offered after patients' rights opponents in the White House exerted pressure on Rep. Norwood to abandon a stronger bill. The legislation would allow a patient to sue a health maintenance organization (HMO) in state court but with federal, not state, law governing. An employer could remove cases to federal court. The bill would limit non-economic damages to $1.5 million. Punitive damages would be limited to the same amount and only allowed when a decision-maker fails to abide by a grant of benefits by an independent medical reviewer. Adopted 218-213. August 2, 1001. A no vote is a +.

(15) **HR 2563. Patients' Bill of Rights.** Passage of the bill to provide federal health care protections, such as access to specialty and emergency room care, and require that health maintenance organizations (HMOs) have an appeals process for patients who are denied care. This weakened legislation was offered to head off consideration of a stronger version. A patient denied care could sue an HMO in state and federal court but first must exhaust internal and external appeals processes. Passed 226-203. August 2, 2001. A no vote is a +.

(16) **HR 2944. Domestic Partner Benefits.** Weldon (R-FL) amendment to the FY 2002 District of Columbia Appropriations Bill that would prohibit the use of local, as well as federal, funds to extend city employees' health benefits to unmarried domestic partners. Rejected 194-226. September 25, 2001. A no vote is a +.

(17) **HR 2586. U.S. Military Personnel Overseas/Abortions.** Sanchez (D-CA) amendment to the FY 2002 Defense Authorization Bill which allows female military personnel stationed at U.S. bases overseas to undergo an abortion at medical facilities there provided they pay for it themselves and a doctor consents to perform the operation. Rejected 199-217. September 25, 2001. A yes vote is a +.

(18) **HR 2975. Anti-Terrorism Authority.** Adoption of the rule to provide for House consideration of the bill that would expand law enforcement's power to investigate suspected terrorists and beef up domestic surveillance. The legislation threatens the civil liberties, civil rights, and due process protections guaranteed individuals in the United States. Adopted 214-208. October 12, 2001. A no vote is a +.

(19) **HR 3090. Economic Stimulus.** Passage of the Republican version of the post- September 11 economic stimulus package. The bill would grant businesses and individuals $99.5 billion in federal tax cuts in fiscal 2002, and a total of $159.4 billion in reductions over 10 years. Additionally, the bill would allow more individuals to receive tax rebates for 2000, accelerate a reduction of the 27 percent tax bracket to 25 percent, lower the capital gains tax rate from 20 percent to 18 percent and eliminate the corporate alternative minimum tax. Also, the legislation would provide $3 billion to states for health insurance for the unemployed. Passed 216-214. October 24, 2001. A no vote is a +.

(20) **HR 3000. Trade Promotion Authority/Fast Track.** Passage of the bill to allow expedited negotiation and implementation of trade agreements between the executive branch and foreign countries. The bill includes provisions requiring increased consultations with Congress on any proposed changes of tariffs for imports of sensitive agriculture products and on trade disparities for textile products. Passed 215-214. December 6, 2001. A no is a +.

Appendix D. Code

```
logcall<-function(data,
          rownames=as.character(1:dim(data)[1]),
          rowlabs=as.character(1:dim(data)[1]),
          ndim=2,
          eps=1e-3,
          imax=10000,
          correct=TRUE,
          extreme=TRUE,
          form="logit",
          offset=1.20) {
name<-deparse(substitute(data))
outfile<-file(paste(name,"out",sep="."),"w")
vlog<-function(a,b) ifelse(b>0,a*log(b),0)
ls<--2*length(which(!is.na(data)))*log(.5)
itel<-1
z<-ifelse(is.na(data),0,-4*(data-.5))
repeat{
          a<-apply(z,2,mean)
          z<-z-a
          sv<-La.svd(z,nu=ndim,nv=ndim,method="dgesdd")
          x<-sv$u; y<-sv$d[1:ndim]*(sv$v)
          aa<-(x%*%y)+a
          pr<-1/(1+exp(aa))
          tb<-table(as.vector(data),as.vector(ifelse(pr
              >.5,1,0)))
          lk<-vlog(data,pr)+vlog(1-data,1-pr)
          lt<--2*sum(lk[which(!is.na(data))])
          cat("Iteration:␣",formatC(itel,digits=6,width=6),
               "␣Deviance:␣",formatC(ls,digits=6,width=12,
                    format="f"),
               "␣==>",formatC(lt,digits=6,width=12,format="f")
                    )
          if (correct)
               cat("␣Correct:␣",formatC(sum(diag(tb))/sum(tb),
                    digits=6,width=10,format="f"))
```

```
          cat("\n")
          if ((abs(lt-ls)<eps) || (itel==imax)) break()
          itel<-itel+1
          ls<-lt
          z<-ifelse(is.na(data),aa,aa-4*(data-pr))
          }
radius<-mean(rowSums(x^2))
xx<-c(min(x[,1]),max(x[,1]))
yy<-c(min(x[,2]),max(x[,2]))
zz<-c(min(xx[1],yy[1]),max(xx[2],yy[2]))
pdf(file=paste(paste(name,"_row",sep=""),"pdf",sep="."),
    encoding="MacRoman")
plot(x,type="n",main=paste("Ideal point plot for",name),
          xlab=paste("dimension",1),ylab=paste("dimension",2)
              ,xlim=offset*xx,ylim=offset*yy)
text(x,rownames,cex=.5,col="red")
dev.off()
pdf(file=paste(paste(name,"_col",sep=""),"pdf",sep="."),
    encoding="MacRoman")
plot(x,type="n",main=paste("Roll call plot for",name),
          xlab=paste("dimension",1),ylab=paste("dimension",2)
              ,xlim=offset*xx,ylim=offset*yy)
text(x,rowlabs,cex=.5,col="green")
for (i in 1:dim(y)[2]) {
     intercept<--a[i]/y[2,i]; slope<--y[1,i]/y[2,i]
          abline(intercept,slope)
          u<-slope*intercept; v<-1+(slope^2); w<-(radius*v)-(
              intercept^2)
          if (w>0) {
               x1<--(u+sqrt(w))/v; x2<--(u-sqrt(w))/v;
                    y1<-intercept+(slope*x1); y2<-intercept+(
                        slope*x2)
                    text(x1,y1,as.character(i),col="red")
                    text(x2,y2,as.character(i),col="red")
                    }
          }
```

```
dev.off()
close(outfile)
list(intercepts=a, rowpoints=x, columnpoints=y, probabilities=
    pr)
}
```

References

Ada. 2001 Voting Record: Shattered Promise of Liberal Progress. *ADA Today*, 57 (1):1–17, March 2002.

U. Böckenholt. Analyzing comparative judgments with and without a scale origin. *Psychological Methods*, in press.

D. Böhning. Multinomial logistic regression algorithm. *Annals of the Institute of Statistical Mathematics*, 44(1):197–200, 1992.

D. Böhning and B.G. Lindsay. Monotonicity of quadratic-approximation algorithms. *Annals of the Institute of Statistical Mathematics*, 40(4):641–663, 1988.

J. Clinton, S. Jackman, and D. Rivers. The statistical analysis of roll call data. URL `http://jackman.stanford.edu/papers/masterideal.pdf`. 2003.

C. Coombs. *A Theory of Data*. Wiley, 1964.

V.A. Daugavet. Variant of the stepped exponential method of finding some of the first characteristics values of a symmetric matrix. *USSR Computation and Mathematical Physics*, 8(1):212–223, 1968.

J. De Leeuw. Block Relaxation Methods in Statistics. In H.H. Bock, W. Lenski, and M.M. Richter, editors, *Information Systems and Data Analysis*, Berlin, 1994. Springer Verlag.

J. De Leeuw. Homogeneity analysis of pavings. URL `http://jackman.stanford.edu/ideal/MeasurementConference/abstracts/homPeig.pdf`. August 2003a.

J. De Leeuw. Least squares optimal scaling of partially observed linear systems. UCLA Statistics Preprints 360, UCLA Department of Statistics, 2003b. URL `http://preprints.stat.ucla.edu/360/`.

J. De Leeuw and G. Michailidis. Block Relaxation Algorithms in Statistics. URL `http://gifi.stat.ucla.edu/pub/block.pdf`. 1999.

G.H. Fischer and I.W. Molenaar. *Rasch Modles. Foundations, Recent Developments, and Applications.* Springer, 1995.

A. Gifi. *Nonlinear multivariate analysis*. Wiley, Chichester, England, 1990.

H.F. Gollob. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, 33:73–116, 1968.

J.C. Gower and D.J. Hand. *Biplots*. Number 54 in Monographs on Statistics and Applied Probability. Chapman & Hall, 1996.

M.J. Greenacre. *Theory and applications of correspondence analysis*. Academic Press, New York, New York, 1984.

W.J. Heiser. Convergent computing by iterative majorization: theory and applications in multidimensional data analysis. In W.J. Krzanowski, editor, *Recent Advanmtages in Descriptive Multivariate Analysis*. Oxford: Clarendon Press, 1995.

J. Hemelrijk. Underlining random variables. *Statistica Neerlandica*, 20:1–7, 1966.

M. Hubert, P.J. Rousseeuw, and S. Verboven. A fast robust method for principal components with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 60:101–111, 2002.

S. Jackman. Estimation and inference are missing data problems: Unifying social science statistics via bayesian simulation. *Political Analysis*, 8(4):307–332, 2000.

S. Jackman. Multidimensional analysis of roll call data via bayesian simulation: Identification, estimation, inference, and model checking. *Political Analysis*, 9 (3):227–241, 2001.

N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Volume I*. Wiley, second edition, 1994.

K. Lange, D.R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, 9:1–20, 2000.

R. R. Meyer. Sufficient conditions for the convergence of monotonic mathematical programming algorithms. *Journal of Computer and System Sciences*, 12:108–121, 1976.

A. M. Ostrowski. *Solution of Equations and Systems of Equations*. Academic Press, New York, N.Y., 1966.

G. Pison, P.J. Rousseeuw, P. Filzmozer, and C. Croux. Robust factor analysis. *Journal of Multivariate Analysis*, 84:145–172, 2003.

K. Poole and H. Rosenthal. *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press, 1997.

K. T. Poole. Nonparametric unfolding of binary choice data. *Political Analysis*, 8 (2):211–237, 1999.

K.T. Poole. The geometry of multidimensional quadratic utility in models of parliamentary roll call voting. *Political Analysis*, 9(3):211–226, 2001.

K.T. Poole and H. Rosenthal. A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29(2):357–384, 1985.

D. Rivers. Identification of multidimensional spatial voting models. URL http://jackman.stanford.edu/ideal/MeasurementConference/abstracts/river03.pdf. 2003.

M.R. Sampford. Some Inequalities on Mill's ratio and Related Functions. *Annals of Mathematical Statistics*, 24:130–132, 1953.

Y. Takane and M. Hunter. Constrained principal components analysis: A comprehensive theory. *Applicable Algebra in Engineering, Communication and Computing*, 12:391–419, 2001.

Y. Takane, H.A.L Kiers, and J. De Leeuw. Component analysis with different sets of constraints on different dimensions. *Psychometrika*, 60:259–280, 1995.

Y. Takane and T. Shibayama. Principal component analysis with external information on both subjects and variables. *Psychometrika*, 56:97–120, 1991.

W. I. Zangwill. *Nonlinear Programming: a Unified Approach*. Prentice-Hall, Englewood-Cliffs, N.J., 1969.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu