Non-Parametric Unfolding of Binary Choice Data

Keith T. Poole
Graduate School of Industrial Administration
Carnegie-Mellon University

26 August 1998

Abstract

This paper shows a general non-parametric unfolding technique for maximizing the correct classification of binary choice or two-category data. The motivation for and the primary focus of the unfolding technique is parliamentary roll call voting data. However, the procedures that implement the unfolding also can be applied to the problem of unfolding rank order data as well as analyzing a data set that would normally be the subject of a probit, logit, or linear probability analysis.

To unfold binary choice data two subproblems must be solved. First, given a set of chooser or legislator points a cutting plane must be found such that it divides the legislators/choosers into two sets that reproduce the actual choices as closely as possible. Second, given a set of cutting planes for the binary choices a point for each chooser or legislator must be found which reproduces the actual choices as closely as possible. Solutions for these two problems are shown in this paper. Monte-Carlo tests of the procedure shows it to be highly accurate in the presence of error and missing data.

1. Introduction

The purpose of this paper is to show a general non-parametric unfolding technique for maximizing the correct classification of binary choice or two-category data. The unfolding technique is non-parametric because the only assumptions made are that the choice space is Euclidean and that individuals making choices behave as if they utilize symmetric, single-peaked preferences. Other than these assumptions, no assumptions are made about the functional form of individuals' preferences and no assumptions are made about the distributional form of individuals' errors in making choices. The motivation for and the primary focus of the unfolding technique is parliamentary roll call voting data but the procedures that implement the unfolding also can be used to solve the problem of unfolding rank order data.

A roll call vote in a legislative assembly using standard parliamentary rules consists of each legislator casting a Yes (Yea) or No (Nay) vote on the motion on the floor. Typically, if the motion fails the status quo prevails. Consequently, each roll call vote can be considered to be a choice between *two* policy outcomes – one corresponding to Yea and one corresponding to Nay – and roll calls can have policy outcomes in common (e.g., an amendment to a bill, an amendment to an amendment, a motion to table, etc.).

Let s denote the number of policy dimensions, which are indexed by k=1,…,s; let p denote the number of legislators (i=1,…,p); and q denote the number of roll call votes (j=1,…,q). Let **X** be the p by s matrix of legislator coordinates, and let **T** be the p by q matrix of observed choices. The choices will simply be Yea or Nay and will be represented as "y" and "n" respectively. **T** can contain missing entries. Because

legislators are assumed to have symmetric single-peaked preferences around their ideal points, if there were no error they would vote for the alternative closest to them in the policy space on any roll call.

In the perfect case, an s-dimensional hyperplane that is both perpendicular to the line joining the Yea and Nay policy points and passes through the midpoint of the Yea and Nay policy points, separates the legislators voting Yea from the legislators voting Nay. The normal vector to this *cutting plane* is parallel to the line joining the Yea and Nay policy points. Note that, in the case of perfect voting, the policy points are not identified – any pair of points on a line perpendicular to the plane that are on opposite sides and equidistant from the plane would produce the same pattern of votes. However, the cutting plane is identified. Hence this is an unfolding which recovers points representing the legislators and *implicitly* pairs of points for each roll call vote albeit only in the form of the cutting planes which pass between the pairs of points.[1]

A plane is defined as $\underline{\mathbf{z}}\,\underline{\mathbf{n}} = \underline{\mathbf{v}}\,\underline{\mathbf{n}}$ where $\underline{\mathbf{z}}$, $\underline{\mathbf{n}}$, and $\underline{\mathbf{v}}$ are s by 1 vectors and the plane consists of all points $\underline{\mathbf{z}}$ such that $(\underline{\mathbf{z}} - \underline{\mathbf{v}})$ is perpendicular to the normal vector, $\underline{\mathbf{n}}$, and $\underline{\mathbf{v}}$ is some point in the plane. In this context, the problem is to solve for the normal vector, $\underline{\mathbf{n}}$. (Given $\underline{\mathbf{n}}$, as shown below, it is simple to find the point on $\underline{\mathbf{n}}$ through which the plane passes.) Let **N** be the q by s matrix of normal vectors for the q cutting planes. Given the number of dimensions, s, the classification problem consists of finding estimates of **X** and **N**, denoted as **X*** and **N*** respectively, which maximize the correct classifications.

In sum, only two basic assumptions are made: 1) the choice space is Euclidean; and 2) the individuals making choices behave as if they utilize symmetric, single-peaked preferences. Consequently, given a matrix of roll calls, the non-parametric unfolding

problem consists of finding a set of legislator points and a set of cutting planes corresponding to each binary choice in an Euclidean space of s dimensions such that each cutting plane divides the legislators into two sets that reproduce the actual choices as closely as possible.

In one dimension this consists of finding a joint rank ordering of the legislators and roll call midpoints (ties are permitted, an example is shown in Table 7 below) that maximizes correct classification. The one-dimensional scaling procedure resembles classical Guttman scaling (see Figure 2). Given a rank order of legislators, the global maximum in classification can be found for every roll call. Similarly, given a rank order of the roll call midpoints, the global maximum in classification can be found for every legislator. The two are symmetric in one dimension.

In two or more dimensions this symmetry disappears. For example, in two dimensions, q cutting lines create a maximum of $q(q-1)/2 + q + 1$ regions (Coombs, 1964, p. 262) with each region corresponding to a voting pattern – e.g., yynnynyn…nn – on the q roll calls. The problem is to place each legislator in a region that best matches the legislator's observed pattern of roll call votes. Given the legislator points, the problem is to find a cutting line for each roll call that divides those legislators voting Yea from those voting Nay such that correct classification is maximized. Solutions for these two problems are shown in the next two sections.

When the number of legislators is 100 or greater and the number of roll calls is on the order of 500 – typical of national legislatures, for example, the U.S. Senate – then the recovery of the legislators and cutting lines is very precise. With 500 roll calls, there are a maximum 125,251 regions in two dimensions and over 20,000,000 in three dimensions.

Most of these regions are so small that a typical legislator's point is very precisely pinned down. In fact the recovery of the legislator coordinates is virtually identical to those recovered by parametric procedures that must make strong assumptions about the interpersonal comparability of individuals' utility and the function form of the error distribution (e.g., Heckman and Snyder, 1997; Poole and Rosenthal, 1997).

The non-parametric unfolding technique developed below can also be used to analyze the general rank ordering problem. Suppose we have the rank orders over a set of stimuli from a group of individuals. For example, for 6 alternatives:

1. A>B>D>F>C>E
2. B>C>D>A>E>F
3. C>A>B>D>F>E

Etc.

p. D>A>C>F>E>B

The individual rank orders can be converted into roll call data by viewing the stimuli as "voting" between pairs of individuals. Namely, for each pair of individuals, let the stimulus vote for the individual who has the stimulus higher in her ordering. For example, consider individuals 1 and 2 shown above. Let "Yea" be a vote for individual 1 and "Nay" be a vote for individual 2. Stimulus A would vote Yea, B Nay, C Nay, D would abstain, E Nay, and F Yea. This produces a q by p(p-1)/2 matrix of "roll calls". The unfolding technique will recover points representing the stimuli and cutting planes between each pair of individuals. The estimates of the stimuli will be very precise using this technique. An example is shown in Section 5.

Sections 2-4 develop the non-parametric unfolding procedure. Section 2 shows a solution for finding the optimal cutting plane given a configuration of legislators, section 3 shows a solution for finding the optimal legislator point given a set of cutting planes, and section 4 shows Monte-Carlo tests of the unfolding procedure. Empirical applications are shown in Section 5.

2. Finding the Optimal Cutting Plane

Given the p by s matrix, **X**, of legislator coordinates and the p by 1 vector of votes on the jth roll call, $\underline{\mathbf{t}}$, the problem is to find the plane that divides the legislators into two groups such that the number of correct classifications is maximized. Figure 1 shows an example in two dimensions.

Figure 1 about Here

Figure 1 illustrates the fact that the cutting plane problem is equivalent to finding a vector – in this case, $\underline{\mathbf{n}}$ – such that when the legislator points are projected onto the vector a *cutting point* can be found that maximizes the correct classifications. By definition, all points in the cutting plane are projected onto this cutting point. The problem has two distinct parts. First, given an estimated normal vector, the plane perpendicular to the normal vector which maximizes correct classifications must be found; and second, given an estimated cutting plane, the orientation of the plane in the space must be changed so that a better estimate of the normal vector is found.

Calculating the Correct Classifications

Let the legislator coordinates lie within the s dimensional unit hypersphere and let the origin of the space be placed at the centroid of the legislator coordinates; that is, let

$$\sum_{k=1}^{s} x_{ik}^2 \leq 1 \quad , i=1,...,p \qquad \text{and} \qquad \sum_{i=1}^{p} x_{ik} = 0 \quad , k=1,...,s$$

In addition, let $\underline{\mathbf{n}}_j$ be the normal vector for the jth roll call that maximizes correct classifications. Without loss of generality $\underline{\mathbf{n}}_j$ can be constrained to be of unit length; i. e., $\underline{\mathbf{n}}_j \underline{\mathbf{n}}_j = 1$. The projections (see Figure 1B) are, therefore:

$$\mathbf{X}\underline{\mathbf{n}}_j = \underline{\mathbf{w}} \qquad (1)$$

Note that the elements in the p-length vector, $\underline{\mathbf{w}}$, range from -1 to +1. The elements in $\underline{\mathbf{w}}$ all lie on a line that passes through the origin of the s-dimensional unit hypersphere in the direction of the normal vector with exit points $-\underline{\mathbf{n}}_j$ and $+\underline{\mathbf{n}}_j$ respectively. Hereafter, this will be referred to as the *projection line*.

Let $\underline{\mathbf{n}}_j^*$ be an estimate of $\underline{\mathbf{n}}_j$ and let $\underline{\mathbf{w}}^*$ be the corresponding estimate of $\underline{\mathbf{w}}$. The correct classifications associated with $\underline{\mathbf{n}}_j^*$ can be calculated quite easily. Figure 2 illustrates the method.

Figure 2 about Here

For ease of exposition, let the projected legislator coordinates from left to right be denoted in order as w_1 to w_p such that $-1 \leq w_1 \leq w_2 \leq w_3 \leq ... \leq w_p \leq +1$ and the "y"s and "n"s above the projection line in the Figure indicate how the corresponding legislators voted on the jth roll call. There are p+1 possible regions that the cutting point could be in -- $(-1, w_1)$, (w_1 , w_2), ... , $(w_p , +1)$ -- and for each region there are exactly 2 possible

perfect voting patterns for an overall total of 2(p+1) possible perfect voting patterns. However, region (w_p , +1) is redundant since it produces the same perfect patterns as the region (-1 , w_1) so it may be discarded leaving 2p unique perfect voting patterns to consider.

Since there are only 2p perfect patterns, it is a simple matter to compare each perfect pattern with the actual pattern of votes, $\underline{\mathbf{t}}_j$. This can be done very efficiently by first assuming that the cutting point is in the region (-1 , w_1) and calculating the corresponding number of correct classifications. Next assume that the cutting point is in the region (w_1 , w_2). Only one calculation has to be made to get the correct classifications for this cutting point since the only change is that the cutting point has been moved from the left of w_2 to the right of w_2. If there is no missing data, either the correct classification increases by 1 or decreases by 1 when the cutting point is moved from the left of w_2 to the right of w_2. Similar reasoning holds for the remaining points. For each possible cutting point the correct classification corresponding to the two possible perfect patterns can be calculated. The estimated cutting point is set equal to the midpoint of the region for which correct classification is a maximum. For the example shown in Figure 2, placing the cutting point at the position of any of the three asterisks would produce only two classification errors for a correct classification of p-2.

Note that this process is equivalent to moving the cutting plane through the unit hypersphere along the estimated normal vector, $\underline{\mathbf{n}}_j^*$.

Calculating the Optimal Normal Vector

Let c* denote the cutting point that maximizes correct classification on the projection line formed by the elements of $\mathbf{X}\underline{\mathbf{n}}_j^* = \underline{\mathbf{w}}^*$. The point c* is therefore:

$$\underline{\mathbf{z}}\,\underline{\mathbf{n}}_j^* = \underline{\mathbf{v}}\,\underline{\mathbf{n}}_j^* = c^* \qquad (2)$$

Given $\underline{\mathbf{n}}_j^*$ and c*, the estimated cutting plane consists of all points $\underline{\mathbf{v}}$ satisfying equation (2). In order to get a new estimate of $\underline{\mathbf{n}}_j$, the estimated cutting plane given by equation (2) must be moved through the space in a direction that increases correct classification. This is accomplished by moving the cutting plane towards the legislator points that are classification errors.

To do this, a matrix is created by projecting all the *correctly classified* legislator points onto the surface of the cutting plane while leaving the incorrectly classified legislators at their original positions. In two dimensions this produces a line through the space made up of correctly classified legislators (the current cutting plane) around which is a scattering of points corresponding to the incorrectly classified legislators (see Figure 3). Specifically, let $\underline{\mathbf{x}}_i$ be the s by 1 vector denoting the ith legislator's point in the space and let w_i be the corresponding point on the projection line from equation (1). Construct a p by s matrix, **V**, as follows: if legislator i is correctly classified, then her point is projected onto the cutting plane and that point becomes the ith row of **V**; if legislator i is incorrectly classified, then her point remains at its original position and that point becomes the ith row of **V**. That is:

$$\begin{aligned} \underline{\mathbf{v}}_i &= \underline{\mathbf{x}}_i + (c^* - w_i)\underline{\mathbf{n}}_j^* \quad \text{if correctly classified} \\ \underline{\mathbf{v}}_i &= \underline{\mathbf{x}}_i \qquad\qquad\qquad\quad \text{if incorrectly classified} \end{aligned} \qquad (3)$$

Figure 3 about Here

Without loss of generality, the centroid of **V** can be placed at the origin. That is, let **m** be the s length vector of the means of **V**, and let J_p be a p by 1 vector of ones. Define **V*** as

$$\mathbf{V^*} = \mathbf{V} - J_p\mathbf{m} \quad (4)$$

Panel A of Figure 3 shows a vote in two dimensions which would be perfectly classified by the indicated cutting line. Panel B shows the **V*** produced by using an initial estimate of $\mathbf{n}_j^*$ = (0 , 1) -- that is, an estimated normal vector perpendicular to the true normal vector. All the "y" and "n" tokens off the plane are classification errors. Clearly, if the plane were moved counter-clockwise towards the errors a better fit would be obtained.

This is accomplished by using the Eckart and Young (1936) lower-rank-matrix approximation theorem. Let the singular value decomposition of **V*** be

$$\mathbf{V^*} = \mathbf{U}\mathbf{L}\mathbf{Q} \quad (5)$$

where **U** is a p by s orthogonal matrix, **Q** is an s by s orthogonal matrix, and **L** is an s by s diagonal matrix containing the singular values in descending order on the diagonal.

By the Eckart-Young theorem, the best fitting line through the scatterplot shown in panel B of Figure 3 is found by inserting a zero in place of the second singular value in **L** and remultiplying. That is, let $\mathbf{L}^{\#}$ be the s by s diagonal matrix identical to **L** except for the replacement of the sth singular value by zero, then the estimated hyperplane is:

$$\mathbf{V}^{\#} = \mathbf{U}\mathbf{L}^{\#}\mathbf{Q} \quad (6)$$

where $\mathbf{V}^{\#}$ will be of rank s-1 by construction.

Let $\underline{\mathbf{n}}_j^{\#}$ be the normal vector of the hyperplane defined by $\mathbf{V}^{\#}$ and let $\underline{\mathbf{q}}_s$ be the sth singular vector of $\mathbf{Q}$. It is straightforward to show that $\underline{\mathbf{n}}_j^{\#} = \underline{\mathbf{q}}_s$. By the definition of a plane:

$$\mathbf{V}^{\#}\underline{\mathbf{n}}_j^{\#} = J_p c^{\#} \qquad (7)$$

where J_p is a p by 1 vector of ones and $c^{\#}$ is a constant. Recall from equation (4) that

$$\sum_{i=1}^{p} v_{ik}^{*} = 0 \quad , k=1,...,s$$

Hence,

$$\sum_{i=1}^{p} v_{ik}^{\#} = 0 \quad , k=1,...,s$$

because the singular vectors of $\mathbf{U}$ in equations (5) and (6) sum to zero. Therefore, by simply adding up all p elements of the vectors on either side of the equality in equation (7) it must be the case that $c^{\#} = 0$. This allows equation (7) to be rewritten as

$$\mathbf{U}\mathbf{L}^{\#}\mathbf{Q}\,\underline{\mathbf{n}}_j^{\#} = \underline{\mathbf{0}}_p \qquad (8)$$

where $\underline{\mathbf{0}}_p$ is a p length vector of zeroes. Let $\mathbf{L}^{\#-1}$ be an s by s diagonal matrix with diagonal entries that are the reciprocals of the non-zero diagonal entries of $\mathbf{L}^{\#}$. Multiplying both sides of equation (8) by $\mathbf{L}^{\#-1}\mathbf{U}$

$$\mathbf{L}^{\#-1}\mathbf{U}\,\mathbf{U}\mathbf{L}^{\#}\mathbf{Q}\,\underline{\mathbf{n}}_j^{\#} = \mathbf{L}^{\#-1}\mathbf{U}\,\underline{\mathbf{0}}_p$$

this reduces to

$$\mathbf{Q}^{*}\,\underline{\mathbf{n}}_j^{\#} = \underline{\mathbf{0}}_s$$

where $\underline{\mathbf{0}}_s$ is an s length vector of zeroes, and $\mathbf{Q}^{*}$ is identical to $\mathbf{Q}$ except the sth column of $\mathbf{Q}^{*}$ is all zeroes (hence, the sth row of $\mathbf{Q}^{*}$ is all zeroes). Now, $\underline{\mathbf{n}}_j^{\#}$ cannot be a vector of zeroes since, by definition, $\underline{\mathbf{n}}_j^{\#}\,\underline{\mathbf{n}}_j^{\#} = 1$. Hence, $\underline{\mathbf{n}}_j^{\#} = \underline{\mathbf{q}}_s$ is a solution for equation (8).

In sum, calculating the optimal $\underline{\mathbf{n}}_j$ consists of the following steps:

1) Obtain a starting estimate of $\underline{\mathbf{n}}_j^*$ using simple OLS (linear probability) or two-group linear discriminate analysis.
2) Calculate the correct classifications associated with $\underline{\mathbf{n}}_j^*$.
3) Construct **V*** using equations (3) and (4).
4) Perform singular value decomposition of **V***, **U**$\mathbf{L}\mathbf{Q}$ ¢
5) Use the sth singular vector of $\mathbf{Q}$, $\underline{\mathbf{q}}_s$, as the new estimate of $\underline{\mathbf{n}}_j$.
6) Go to (2).

In a perfect case like that shown in Figure 3, this cutting plane procedure will almost always quickly iterate into the true cutting plane. Two group linear discriminant analysis produces a very good starting estimate for $\underline{\mathbf{n}}_j^*$ -- correct classification is almost always about 90 percent (OLS yields nearly identical starting estimates). With perfect data, the rate of convergence is a function of the number of errors. As the number of errors decreases, the mass of the correctly classified choices increases thereby producing very small changes in the newly estimated normal vectors. The procedure is stopped when the sum of squared differences in $\underline{\mathbf{n}}_j^*$ divided by s changes less than .0001 between iterations.

Table 1A shows a Monte-Carlo study of the cutting plane procedure using perfect data for 100 legislators and 500 roll calls for 2 through 10 dimensions and Table 1B shows a Monte-Carlo study of contaminated data in 1 to 3 dimensions with varying error levels and types of error. Results for one dimension are not shown in Table 1A since correct classification will always be 100% if perfect data is used. The 100 legislators and 500 pairs of policy points were randomly drawn from a uniform distribution through the unit

hypersphere. The policy points were randomly drawn but in such a way so as to produce an average majority margin of about 67 percent (typical of U. S. congressional roll call data – see Table 6). A maximum of 50 iterations through steps (2) - (5) above were allowed.

Table 1 about Here

The cutting plane procedure performs very well. The number of dimensions does not appear to play any role in the accuracy of the procedure. For example, for the ten trials in 10 dimensions, the 5000 total estimated $\underline{\mathbf{n}}_j^*$'s correctly classified 499,936 of 500,000 choices (99.99 percent). With 100 legislators the recovery of the true normal vectors is quite good. In two dimensions, the average of the cosines computed between the true normal vectors and the estimated normal vectors is .998 for all roll calls and .999 for roll calls with at least 10 percent in the minority (90-10 or better). This average cosine falls as the number of dimensions is increased because of the simple fact that there is more space between the legislators so that the cutting planes can be moved slightly without affecting the classification, but even at ten dimensions the recovery is quite good. Note that, in four or less dimensions – the dimensionality of most practical applications -- the recovery is almost exact.

Error was introduced by making the legislator choices probabilistic such that the further a legislator is from the cutting plane, the less likely the legislator will make a voting error. Specifically, an indirect utility function for each legislator was created -- $u_{ij.} + \varepsilon_{ij.}$ – where $u_{ij.}$ is the deterministic portion of the utility function and $\varepsilon_{ij.}$ is the stochastic portion. The deterministic portion is assumed to be an exponential function of the

negative of the squared distance from the legislator to the "y" and "n" alternatives and ε_{ijy} and ε_{ijn} were drawn from the Normal, Uniform, and Logit distributions, respectively.

Table 1B shows that the procedure does a good job correctly classifying the true roll call choices and recovering the true normal vectors – especially at the 15 percent error level which is the approximate level of the error found in the U.S. Congressional roll call data.[2] Finally, as one would expect, increasing the number of legislators increases the accuracy of the recovery.

When error is present the cutting plane procedure converges very quickly. An example is shown in Figure 4 that uses the same configuration of legislator ideal points as Figure 3. The choices of 78 of the 435 legislators have been modified so that they are "errors" – "N's" on the "Y" side of the true cutting line and "Y's" on the "N" side of the true cutting line. The cutting plane procedure converges on the 30th iteration as shown in Panel D. As shown by Panels B and C, in the error case the converged cutting plane may not be the one that maximizes classification – however, it will invariably be very close to the optimal cutting plane. This is easily dealt with by simply storing the iteration record and using the normal vector corresponding to the best classification. This works very well in practice.

Figure 4 about Here

Relationship With Probit and Manski's Maximum Score Estimator

Given a simple two category dependent variable and a set of fixed independent variables, the cutting plane procedure can be used to estimate a vector of coefficients for the independent variables that maximizes correct classification of the dependent variable.

In this instance, with the independent variables scaled so as to lie within a unit hypersphere, the normal vector, $\underline{\mathbf{n}}_j^*$, produced by the cutting plane procedure, plays the role of the coefficient vector, $\underline{\mathbf{b}}$, in a standard Probit, Logit, or linear probability analysis. For example, in a Probit analysis, if the estimated **b**'s for the independent variables, $\mathbf{b}_1$, $\mathbf{b}_2$, … , $\mathbf{b}_s$, are normalized so that their sum of squares is equal to one, then they constitute a normal vector to a plane upon which the choice probabilities are exactly .5/.5. That is, in terms of Figure 1, all the points on the cutting plane have choice probabilities of .5/.5 and are projected onto the projection line at the cutting point, **c***, which is determined by the intercept term, $\mathbf{b}_0$. In this context, the normal vector is the direction of maximum increase/decrease in probability and the $\mathbf{1/s}$ in the standard Probit expression determines how rapidly the probability rises/falls along the normal vector from the cutting plane to the rim of the hypersphere; that is from **c*** to $-\underline{\mathbf{n}}_j$ and $+\underline{\mathbf{n}}_j$.

The cutting plane procedure can also be viewed as a method of implementing Manski's Maximum Score Estimator (see Manski, 1975, 1985; Manski and Thompson, 1986). In its simplest form the Maximum Score Estimator chooses $\underline{\mathbf{b}}$ to maximize correct classification. Part of the MSE process is very similar to the part of the cutting plane procedure illustrated by Figure 2. Namely, given a direction, it is easy to find the classification maximum (Manski and Thompson, 1986, pp. 89-90). However, the MSE approach has no method other than exhaustive search to find a better direction through the space.

Intuitively, in terms of the notation developed above, the MSE algorithm consists of two phases. First, given $\underline{\mathbf{n}}_j^*$, the cutpoint, **c*** is found. Second, let $\underline{\mathbf{n}}_j^1$, $\underline{\mathbf{n}}_j^2$, $\underline{\mathbf{n}}_j^3$, …, $\underline{\mathbf{n}}_j^{s-1}$, be a set of normal vectors orthogonal to $\underline{\mathbf{n}}_j^*$, the algorithm then searches along

these orthogonal vectors for a better solution than the current solution. Unlike the cutting plane procedure that uses equations (3) and (6) to arrive at a better solution for the normal vector, the MSE algorithm has no systematic criterion for selecting a better search direction (Greene, 1993, pp. 658-659).[3]

3. Finding the Optimal Legislator Coordinates

Given the q by s matrix, **N**, of normal vectors and the q by 1 vector of votes of the ith legislator, $\underline{\mathbf{t}}_i$, the problem is to find the legislator point, $\underline{\mathbf{x}}_i$, which maximizes the correct classification. Figure 5 shows an example in two dimensions.

Figure 5 about Here

Figure 5 shows five cutting lines indicated by the numbering at the rim of the circle. The "Y" and "N" on either side of each cutting line indicates how a legislator on that side of the cutting line should vote – "yea" or "nay" respectively. The maximum number of regions created by five cutting lines in two dimensions is 16 and each of these 16 regions can be characterized by a unique vector of votes. Figure 5 only shows 13 regions to emphasize the practical issue that in real world data not all of the theoretically possible regions will be present. For example, cutting lines 2 and 5 intersect outside the circle so that a legislator who voted "y" on both cannot be placed inside the circle without a classification error. In practice, the restriction that the legislator points lie within a unit hypersphere does not pose a problem since the legislator points and cutting planes are iteratively adjusted to maximize correct classification. In this case, if the data were perfect

and two dimensional, the procedure would move the legislator points in such a way that lines 2 and 5 would intersect inside the circle (see Table 4).

Given a legislator's pattern of votes, in this case nnnyn (technically, $\underline{\mathbf{t}}_i$ ¢= [nnnyn]), the problem is to find the region in Figure 5 that maximizes the correct classification. In this example the point "C" is located in the region corresponding to perfect classification. Suppose the initial estimate of the legislator's coordinates is at the origin, point "A" in the Figure. This initial estimate is very poor as it only correctly classifies one of the five votes. The problem is to move the point representing the legislator in a direction that increases the number of correct classifications.

Below a method is shown for finding the maximum classification point along any arbitrary line passing through the space. This method is used to move the legislator point through the space in a city-block fashion by searching along a line parallel to the first dimension and then solving for the point along this line that maximizes classification. Then the legislator point is moved along a line through this new point but parallel to the second dimension. This is done for each dimension in turn and can be repeated as many times as desired. This always converges to a point for which the coordinates are at a *local maximum* in terms of classification. That is, the point cannot be moved parallel to any dimension and have the correct classifications increase.

Let $\underline{\mathbf{x}}_i^{(h)}$ be the initial estimate for legislator i where "h" is the iteration number (1, 2, 3, etc.) and let $\underline{\mathbf{x}}_i^{(a)}$ be a second point. The problem is to find a new estimate, $\underline{\mathbf{x}}_i^{(h+1)}$, on the line passing through $\underline{\mathbf{x}}_i^{(h)}$ and $\underline{\mathbf{x}}_i^{(a)}$ which increases correct classification. Using equation (1), the projection of $\underline{\mathbf{x}}_i^{(h)}$ onto the jth normal vector is:

$$\underline{\mathbf{x}}_i^{(h)\prime}\underline{\mathbf{n}}_j = w_{ij}^{(h)} \qquad (9)$$

Similarly, the projection of the second point onto the jth normal vector is $w_{ij}^{(a)}$. These projections correspond to a correct classification on roll call j depending upon which side of the cutpoint, c_j, they fall. There are six possible orderings of $w_{ij}^{(h)}$, $w_{ij}^{(a)}$, and c_j. For each ordering there are two possible classification outcomes for a total of 12 cases. Table 2 shows each case.

Table 2 about Here

For example, in case 1 both $\underline{\mathbf{x}}_i^{(h)}$ and $\underline{\mathbf{x}}_i^{(a)}$ project to the right of c_j and are on the correct side of the cutting plane for the jth roll call and are therefore correctly classified. Case 2 is the same geometrically only now $\underline{\mathbf{x}}_i^{(h)}$ and $\underline{\mathbf{x}}_i^{(a)}$ are on the wrong side of the cutting plane and are therefore projected as classification errors. Cases 1 to 8 represent no change in classification from moving the legislator point from $\underline{\mathbf{x}}_i^{(h)}$ to $\underline{\mathbf{x}}_i^{(a)}$. For $\underline{\mathbf{x}}_i^{(a)}$ to be an improvement over $\underline{\mathbf{x}}_i^{(h)}$, the number of cases 10 and 12 must be greater than the number of cases 9 and 11.

Consider the effect of moving $\underline{\mathbf{x}}_i^{(a)}$ further from $\underline{\mathbf{x}}_i^{(h)}$. This has no effect on cases 1, 2, and 7 - 12. Only those cases where $\underline{\mathbf{x}}_i^{(a)}$ is between $\underline{\mathbf{x}}_i^{(h)}$ and c_j – cases 3, 4, 5, and 6 – are affected. Depending upon how far $\underline{\mathbf{x}}_i^{(a)}$ is moved away from $\underline{\mathbf{x}}_i^{(h)}$, case 3 could change to case 11 increasing the error by one, case 5 could change to case 9 also increasing the error by one, case 4 could change to case 12 decreasing the error by one, and case 6 could change to case 10 also decreasing the error by one. A similar analysis of the effect of moving $\underline{\mathbf{x}}_i^{(a)}$ towards $\underline{\mathbf{x}}_i^{(h)}$ can also be done.

More generally, consider the line equation:

$$\underline{\mathbf{x}}_i^{(h+1)} = \underline{\mathbf{x}}_i^{(h)} + \alpha(\underline{\mathbf{x}}_i^{(a)} - \underline{\mathbf{x}}_i^{(h)}) \qquad (10)$$

which, when projected onto the jth normal vector, becomes:

$$w_{ij}^{(h+1)} = w_{ij}^{(h)} + \alpha(w_{ij}^{(a)} - w_{ij}^{(h)}) \qquad (11)$$

For a single roll call, it is easy to solve for α; these are shown in Table 2 for all 12 cases. For example, for case 2, α must be chosen so that the projection of $\underline{\mathbf{x}}_i^{(h+1)}$, $w_{ij}^{(h+1)}$, is in the region $(-1, c_j)$.

Given $\underline{\mathbf{x}}_i^{(h)}$ and $\underline{\mathbf{x}}_i^{(a)}$, Table 2 can be used to find the limits of α for each roll call. Let the upper and lower limits for the jth roll call be U_{ij} and L_{ij} respectively. The correct classification associated with $\underline{\mathbf{x}}_i^{(h)}$ can be obtained by setting α=0 and counting the number of roll calls for which $0 \in (L_{ij} , U_{ij})$. Similarly, the correct classification associated with $\underline{\mathbf{x}}_i^{(a)}$ is obtained by setting α=1 and counting the number of roll calls for which $1 \in (L_{ij} , U_{ij})$. In general, define

$$\delta_{ij} = 1 \text{ if } \alpha \in (L_{ij} , U_{ij})$$

$$\delta_{ij} = 0 \text{ if } \alpha \notin (L_{ij} , U_{ij})$$

and the correct classification is simply

$$\delta(\alpha) = \sum_{j=1}^{q} d_{ij} \qquad (12)$$

The α that maximizes $\delta(\alpha)$, the number of correct classifications, can be calculated in a simple manner. First, compute the L_{ij} and U_{ij} for each roll call. Second, rank order the L_{ij} and U_{ij} and use the classification algorithm described in section 2 to calculate the optimal α. Here the L_{ij} play the role of "y" and the U_{ij} play the role of "n". For example, if there exists an α that results in perfect classification, the ordering of L's

and U's will look like (dropping the i subscript to reduce clutter and numbering left to right for convenience):

$$L_1 < L_2 < L_3 < ... < L_q < U_1 < U_2 < U_3 < ... < U_q$$

that is, all the L_j will be less than all the U_j. In this example, perfect classification, $\delta(\alpha) = q$, results from $\alpha \in (L_q , U_1)$.

For example, using the configuration shown in Figure 5, the starting estimate (h=1) $\underline{\mathbf{x}}_i^{(1)}$, is placed at the origin – point "A" – and the second point, $\underline{\mathbf{x}}_i^{(a)}$, is placed just to the right of $\underline{\mathbf{x}}_i^{(1)}$. The resulting rank order of the upper and lower limits is:

$$L_1 < L_5 < L_3 < L_2 < U_3 < U_2 < U_1 < U_5 < L_4 < U_4$$

The rank ordering is almost a perfect pattern in that 4 of the lower limits are below the 5 upper limits; only L_4 is wrongly placed producing one classification error. Consequently, the point resulting from using $\alpha \in (L_2 , U_3)$, $\underline{\mathbf{x}}_i^{(2)}$, point "B" in Figure 5, only has one classification error with 4 correct classifications. (In practice, α is set equal to the midpoint; in this case,$(L_2 + U_3)/2$.) Note that in Figure 5 point "B" is on the wrong side of the cutting line for roll call 4 in the region associated with the pattern nnnnn.

For the second iteration, h=2, the starting estimate is $\underline{\mathbf{x}}_i^{(2)}$ and the second point, $\underline{\mathbf{x}}_i^{(a)}$, is placed just below $\underline{\mathbf{x}}_i^{(2)}$ so that the resulting line is parallel to the second dimension. This produces the rank ordering:

$$L_4 < L_2 < L_5 < L_1 < L_3 < U_4 < U_1 < U_3 < U_5 < U_2$$

The rank ordering is now a perfect pattern with all 5 lower limits below the 5 upper limits so that there are no classification errors. The point resulting from using $\alpha \in (L_3 , U_4)$, $\underline{\mathbf{x}}_i^{(3)}$, point "C" in Figure 5, has 5 correct classifications and no classification error.

The search for the optimal $\underline{\mathbf{x}}_i$ is conducted in a city-block manner. If the starting point is placed at the origin, then in the first iteration the search is along a line through the origin with all but the first dimension coordinates in $\underline{\mathbf{x}}_i^{(1)}$ and $\underline{\mathbf{x}}_i^{(a)}$ set to zero. In the second iteration, the first dimension coordinates are all set equal to the value corresponding to the optimal first dimension value and the 3rd, 4th, …, sth dimensional coordinates in $\underline{\mathbf{x}}_i^{(2)}$ and $\underline{\mathbf{x}}_i^{(a)}$ remain at their original values of zero. The search is along the corresponding line through $\underline{\mathbf{x}}_i^{(2)}$ and $\underline{\mathbf{x}}_i^{(a)}$ which is orthogonal to the first dimension. In the third iteration, the first and second dimension coordinates are set equal to the optimal values from the first and second iterations respectively, and the 4th, 5th, …, sth dimensional coordinates in $\underline{\mathbf{x}}_i^{(3)}$ and $\underline{\mathbf{x}}_i^{(a)}$ remain at their original values of zero. The search is along the corresponding line through $\underline{\mathbf{x}}_i^{(3)}$ and $\underline{\mathbf{x}}_i^{(a)}$ which is orthogonal to the second dimension. This process continues in the same fashion through the sth dimension. Since the search for the optimal $\underline{\mathbf{x}}_i$ is being done city-block-wise, dimensions 1 to s can now be searched again.

In sum, calculating the optimal $\underline{\mathbf{x}}_i$ consists of the following steps:

1) Obtain a realistic starting estimate, $\underline{\mathbf{x}}_i^{(1)}$ (or set $\underline{\mathbf{x}}_i^{(1)}$ equal to the origin, that is, $\underline{\mathbf{x}}_i^{(1)} = \underline{\mathbf{0}}$).

2) Set $\underline{\mathbf{x}}_i^{(a)\prime} = (0.01, x_{i2}^{(1)}, x_{i3}^{(1)}, x_{i4}^{(1)}, x_{i5}^{(1)}, \ldots, x_{is}^{(1)})$, find optimal α and $\underline{\mathbf{x}}_i^{(2)} = \underline{\mathbf{x}}_i^{(1)} + \alpha(\underline{\mathbf{x}}_i^{(a)} - \underline{\mathbf{x}}_i^{(1)})$.

3) Set $\underline{\mathbf{x}}_i^{(a)\prime} = (x_{i1}^{(2)}, 0.01, x_{i3}^{(1)}, x_{i4}^{(1)}, x_{i5}^{(1)}, \ldots, x_{is}^{(1)})$, find optimal α and $\underline{\mathbf{x}}_i^{(3)} = \underline{\mathbf{x}}_i^{(2)} + \alpha(\underline{\mathbf{x}}_i^{(a)} - \underline{\mathbf{x}}_i^{(2)})$.

4) Set $\underline{\mathbf{x}}_i^{(a)\prime} = (x_{i1}^{(2)}, x_{i2}^{(3)}, 0.01, x_{i4}^{(1)}, x_{i5}^{(1)}, \ldots, x_{is}^{(1)})$, find optimal α and $\underline{\mathbf{x}}_i^{(4)} = \underline{\mathbf{x}}_i^{(3)} + \alpha(\underline{\mathbf{x}}_i^{(a)} - \underline{\mathbf{x}}_i^{(3)})$.

5) Set $\underline{\mathbf{x}}_i^{(a)\prime} = (x_{i1}^{(2)}, x_{i2}^{(3)}, x_{i3}^{(4)}, 0.01, x_{i5}^{(1)}, \ldots, x_{is}^{(1)})$, find optimal α and $\underline{\mathbf{x}}_i^{(5)} = \underline{\mathbf{x}}_i^{(4)} + \alpha(\underline{\mathbf{x}}_i^{(a)} - \underline{\mathbf{x}}_i^{(4)})$.

etc.

s+1) Set $\underline{\mathbf{x}}_i^{(a)\prime} = (x_{i1}^{(2)}, x_{i2}^{(3)}, x_{i3}^{(4)}, x_{i4}^{(5)}, \ldots, x_{is-1}^{(s)}, 0.01)$, find optimal α and $\underline{\mathbf{x}}_i^{(s+1)} = \underline{\mathbf{x}}_i^{(s)} + \alpha(\underline{\mathbf{x}}_i^{(a)} - \underline{\mathbf{x}}_i^{(s)})$.

s+2) Go to (2).

Note that classification error *can never increase from one step to the next*. This is true because setting $\alpha = 0$ preserves the current value of classification. This process converges very quickly (usually less than 10 iterations through steps 2 to s+1 above) to a vector of coordinates which is a local maximum in terms of classification. That is, it converges to a point such that $\alpha = 0$ for all s dimensions.

In practice, the starting estimate, $\underline{\mathbf{x}}_i^{(1)}$, and the second point, $\underline{\mathbf{x}}_i^{(a)}$, could be placed anywhere within the s dimensional unit hypersphere. In practical applications the starting estimate is not the origin; rather, realistic starting estimates for the $\underline{\mathbf{x}}_i^{(1)}$'s are obtained from an eigenvector/eigenvector decomposition of the double-centered[4] agreement score matrix computed between legislators. The first s eigenvectors normalized to lie in the unit hypersphere are used as the starts.

If the line through $\underline{\mathbf{x}}_i^{(h)}$ and $\underline{\mathbf{x}}_i^{(a)}$ is parallel to a cutting line then the corresponding difference between $w_{ij}^{(a)}$ and $w_{ij}^{(h)}$, $w_{ij}^{(a)} - w_{ij}^{(h)}$, which is used in Table 2 to find α_j, may be equal to zero. This presents no problem since if the line through $\underline{\mathbf{x}}_i^{(h)}$ and $\underline{\mathbf{x}}_i^{(a)}$ is parallel to a cutting line then the classification on that roll call is the same no matter where on the line $\underline{\mathbf{x}}_i^{(h+1)}$ is located. Consequently, the roll call is not used to locate $\underline{\mathbf{x}}_i^{(h+1)}$. In addition, if the line through $w_{ij}^{(a)}$ and $w_{ij}^{(h)}$ goes through the hypersphere so that it never intersects a

cutting plane this can result in a value of α_j that produces a point that lies outside the unit hypersphere. This is easily handled by computing the upper and lower feasible limits of $\underline{\mathbf{x}}_i^{(h+1)}$ – that is, the values corresponding to the two exit points of the line from the unit hypersphere – and discarding all the L_{ij} and U_{ij} outside the hypersphere. This requires some bookkeeping but it has no effect on the search process. Finally, the search process does not have to be done by moving orthogonally (i.e., city-block-wise) through the hypersphere. However, considerable experimentation shows that it is the most efficient way to proceed.

To guard against bad local maxima (α=0 in s orthogonal directions), multiple starting points for the $\underline{\mathbf{x}}_i^{(1)}$'s are utilized. If different solutions are found (which are rare and almost always close together) then the lines joining the unique local maxima are searched for the best solution. After considerable experimentation, 3 starting points were found to work very well in practice. One starting point is from the eigenvalue-eigenvector decomposition of the double-centered agreement score matrix and the other two are randomly generated. After the first iteration, the legislator point from the previous iteration is used as one of the starting points.

Table 3A shows a Monte-Carlo study of the legislator procedure using perfect data – the true cutting planes are known -- for 100 legislators and 500 roll calls in 2 through 10 dimensions and Table 3B shows results for contaminated data following the same design as that shown in Table 1B. To make the test reasonably stringent, only "unreasonable" starting points are used – namely, the origin and two randomly generated points. Results for one dimension are not shown since classification will always be 100% if perfect data is used. The 100 legislators and pairs of policy points were randomly drawn from a uniform

distribution through the unit hypersphere. The pairs of policy points were drawn in such a way so as to produce cutting lines with an average majority margin of about 67 percent (typical of U.S. congressional roll call data – see Table 6). A maximum of 25 iterations through steps (2) - (s+1) above were allowed.

Table 3 about Here

The legislator procedure works very well – especially at 7 dimensions and below. There is some deterioration in accuracy at 10 dimensions but it still only makes an average of about 43 misclassifications out of 50,000 total choices. For 5 dimensions and below it is practically perfect. Table 3 also shows the average Pearson squared correlations between the true and reproduced legislator coordinates. The average of the worst and best r-squares for the s dimensions are shown.

These r-squares are very high. Even though the legislator procedure is non-parametric, with 500 roll call cutting planes, the unit hypersphere is chopped up into enough regions that, in effect, metric (i.e., ratio scale) information is being extracted from the roll call matrix. In three dimensions with 500 roll calls, there is a theoretical maximum of 20,833,751 regions created by the 500 cutting planes. Obviously, even if only a fraction of these regions are present, their average volume must be very small.

Table 3B is organized in the same fashion as Table 1B. Not surprisingly, as the number of cutting planes increases with the error level held fixed, the precision of the recovery of the legislators increases dramatically. Even at the very high error level of 25

percent, with 500 roll calls in two or three dimensions the recovery of the legislator coordinates is very good.

4. Non-Parametric Unfolding of Binary Choice Matrices

The non-parametric unfolding algorithm consists of three phases:

1) Generate starting values for **X**, **X***, from an eigenvalue/eigenvector decomposition of the legislator by legislator agreement score matrix.
2) Given **X***, find the optimal estimate of **N**, **N***.
3) Given **N***, find the optimal **X***.
4) Go to (2).

Table 4 shows a Monte-Carlo study of the non-parametric unfolding algorithm using perfect data for 100 legislators and 500 roll calls in 1 through 10 dimensions. Only roll calls with margins of 97-3 to 50-50 were used because unanimous and near-unanimous roll calls trivially inflate the number of correct classifications. A maximum of 25 iterations through steps (2) and (3) above were allowed.

Table 4 about Here

The algorithm works well regardless of the number of dimensions. The worst result is for two dimensions where, on average, about 23 of 50,000 choices were misclassified. The accuracy of the recovery of the true configuration of legislators and the true normal vectors declines after 3 dimensions but not very substantially. Even at ten dimensions the average worst Pearson r-square between the true and reproduced legislator coordinates is .943. For four and fewer dimensions, the recovery is very precise.

The algorithm also works reasonably well when the dimensions are not equally salient. For example, in two dimensions if 85 percent of the cutting lines are nearly parallel to the second dimension, the legislator configuration is recovered with reasonable precision – the average r-squares for the first and second dimensions are .94 and .89, respectively. However, in real world applications where noise is present, such data will look like it fits a one-dimension model. Consequently, there is no substitute for the researcher's substantive understanding of the data.

Given the history of other multidimensional scaling techniques, most empirical applications of the non-parametric unfolding technique shown here will be to data matrices with missing entries and the estimated configurations will be in three or fewer dimensions. Missing data presents no problem for the algorithm. In the cutting plane procedure it simply means that the total number of legislators may vary from vote to vote. In the legislator procedure it simply means that the number of cutting lines may vary from legislator to legislator. Handling missing data requires a little bookkeeping but it has no effect on the algorithm.

Table 5 shows a set of experiments with and without error at various levels of missing data. Configurations of 100 legislators and 500 roll calls in 2 and 3 dimensions were randomly generated in the same fashion as those used in the Monte-Carlo experiments shown in Table 4. Error was introduced into the choices by making them probabilistic using the same method as in Table 3B. An error level of about 20 percent was chosen because it is somewhat above the approximate level of error in U.S. congressional roll call data (Poole and Rosenthal, 1997). Matrix entries were randomly removed and the remaining entries were then analyzed by the algorithm in one through

five dimensions. The upper part of Table 5 shows two-dimensional experiments at four different levels of missing data with and without error, and the lower part shows three-dimensional experiments. Each randomly produced matrix was analyzed at each level of missing data so that the same 10 matrices for two or three dimensions (with varying levels of missing entries) are being averaged in each row of the upper or lower parts of the Table.

Table 5 about Here

The accuracy of the recovery of the legislator configuration is quite good and only begins to fall off at 70 percent missing entries. With perfect data the procedure unambiguously finds the true dimensionality. With error there are clear "elbows" at the true dimensionality. The tendency for the correct classification to increase with the percentage of missing data is due to the fact that with more missing data there are fewer roll call cutting planes and hence a legislator's position is not as constrained as it is with complete data. Indeed, the average largest distance to a boundary increases with the level of missing data. This tends to increase the correct classification and decrease the correlation between the true and reproduced legislator configurations. In any event, the results shown in Table 5 suggest that the algorithm will perform well with real world data at realistic levels of missing entries. In particular, with 20 percent missing data there is no appreciable deterioration in performance.

5. Empirical Examples

Non-Parametric Unfolding of U. S. Senate Roll Call Data

Roll call voting in the U.S. Congress has been extensively analyzed by researchers using a wide variety of techniques. This will facilitate the interpretation of the non-parametric unfolding results. Two-dimensional senator coordinates from the non-parametric unfoldings will be compared with those produced by KYST, a multi-dimensional scaling program developed by Kruskal, Young, and Seery (1973), and NOMINATE, a maximum likelihood procedure developed by Poole and Rosenthal (1997).

Table 6 reports the classification results for Senates 80 to 104 in one and two dimensions for the non-parametric procedure. These percentages are about 3 to 5 percentage points better than NOMINATE in both one and two dimensions (Poole and Rosenthal, 1997, chapter 3). This is not surprising given that the NOMINATE procedure maximizes a likelihood function and does not attempt to maximize correct classifications.

Table 6 about Here

Table 6 also shows the squared Pearson correlations between the estimated dimensions of the non-parametric procedure and those produced by KYST and NOMINATE, respectively, in two dimensions. The non-parametric configuration was rotated to best match the NOMINATE and KYST configurations using Schonemann's (1966) technique. These r-squares are, for the most part, very high – most of the first dimension r-squares are above .95 and the second dimension r-squares are mostly above

.9. R-squares were also computed for the Heckman-Snyder configurations. These were nearly the same as those reported for W-NOMINATE because the Heckman-Snyder and NOMINATE configurations are highly correlated (Poole and Rosenthal, 1997, Appendix B). For example, the r-squares between the Heckman-Snyder configuration and the non-parametric configuration for the 85th Senate (shown in Figure 6) are .973 and .944, respectively.

Table 7 shows the estimated rank order from the one-dimensional scaling of the 104th Senate. The ordering is from most liberal (1) to most conservative (103) and it correctly classifies 90.0 percent of the choices (70,976 of 78,882). Campbell of Colorado switched from Democrat to Republican in April of 1995 so he appears twice (ranks 48 and 55). If two or more senators tied in the ranking, the average of the associated ranks was used. For example, 85 senators were more liberal and 15 more conservative than the threesome Mack (R-FL), Coverdell (R-GA), and Coats (R-IN), who were tied. Consequently they all were assigned the average rank of 87.

Table 7 about Here

The polarization of American politics (Poole and Rosenthal, 1997; King, 1998) is evident from an inspection of the table. There is no overlap of the two parties. Campbell's voting record as a Democrat made him the most conservative Democrat in the Senate. His conversion only moved him from 48th to 55th rank – from the right edge of the Democratic party to the midst of the moderates of the Republican party.

Figure 6 shows the two dimensional configuration of senators for the 85th Senate along with a histogram of the roll call cutting line angles. The correct classification was

89.5 percent (20,679 of 23,097). The two major parties are clearly separated with the Democratic Party being split into its Northern and Southern wings. The 85th Senate occurred during the height of the three-party system that lasted from the late 1930s to the late 1970s (Cox and McCubbins, 1993; Poole and Rosenthal, 1997). The approximate angle of a party-line vote and the approximate angle of a conservative coalition vote (Northern Democrats versus a coalition of Southern Democrats and Republicans) are indicated in the histogram of the cutting line angles. The second dimension picked up the split in the Democratic Party over race-related issues.

Figure 6 about Here

Table 8 shows bootstrapped standard errors for the 98 Senators shown in Figure 6. There were only 255 roll calls in the 85th Senate with minority percentages of 2.5 percent or better. From these, 100 samples of 255 roll calls were drawn, with replacement. The non-parametric procedure was run on each of the 100 samples and the standard deviation of the 100 estimates for each Senator for each dimension was computed.

The results are quite good for the first dimension. Eighty of 98 Senators have standard deviations of less than .10. The standard deviations for the second dimension are larger reflecting the fact that the bulk of the cutting lines are between 60 and 120 degrees (see Figure 6). Even so, 72 of 98 Senators have standard deviations of less than .15, which is small relative to the 2-unit diameter of the space. The standard deviations tend to be larger for those Senators near the rim of the space.

Table 8 about Here

Analyzing Rank Order Data: The 1968 U.S. National Election Study Feeling Thermometers

A feeling thermometer measures how warm or cold a person feels towards the stimulus and the measure ranges from 0 – very cold and unfavorable opinion – to 100 – very warm and favorable opinion with 50 being a neutral point. In 1968 respondents were asked to give feeling thermometer ratings to 12 political figures: George Wallace, Hubert Humphrey, Richard Nixon, Eugene McCarthy, Ronald Reagan, Nelson Rockefeller, Lyndon Johnson, George Romney, Robert Kennedy, Edmund Muskie, Spiro Agnew, and Curtis LeMay. A respondent's thermometer scores can be easily converted to simple rank orders. For example, respondent number 4 had the following thermometer scores in the order of the political figures listed above:

0, 0, 50, 0, 60, 60, 70, 50, 97, 60, 60, 50

which produces the following rank ordering from low to high

2, 2, 5, 2, 8.5, 8.5, 11, 5, 12, 8.5, 8.5, 5

Assume that each of the political figures are "legislators" and each unique pair of respondents is a pair of policy points. The "legislators" then vote "yes" for that respondent who has the "legislator" highest in her rank ordering. If the ranks are the same the "vote" is treated as missing data. This creates a roll call matrix where the legislator points are the political figures and the cutting lines correspond to pairs of respondents. Since the number of pairs of respondents is about 50,000, the legislator points can be

recovered with very high precision. Note that this can be accomplished without making the assumption of interpersonal comparison of utility of the respondents.

Because of computer memory limitations, 1200 pairs of respondents were randomly drawn from the data set. Only respondents giving a complete set of thermometer responses were used. This produced a 12 by 1200 matrix of "roll calls". Figure 7 shows the estimated two dimensional configuration.

Figure 7 about Here

These data have been analyzed by variety of scaling techniques (Weisberg and Rusk, 1970; Wang, Schonemann, and Rusk, 1975; Rabinowitz, 1976; Cahoon, Hinich, and Ordeshook, 1978; Poole and Rosenthal, 1984). The configuration shown in Figure 7 is essentially the same as recovered by the other methods. The correct classification was 84.7 percent (11,336 of 13,379). The locations for the political figures are precisely estimated. The maximum distance to a boundary found for any of the 12 points was 0.002.

7. Conclusion

This paper shows a general non-parametric technique for maximizing the correct classification of binary choice data. The motivation for and the primary focus of the unfolding technique is parliamentary roll call voting data but the procedures that implement the unfolding can also be applied to a variety of other problems.

Although neither the cutting plane nor the legislative procedure can formally be shown to converge to the global classification maximum, Monte-Carlo tests show that

both in fact work very well in practice. In the presence of error, because of the way that the cutting plane procedure is operationalized, it almost certainly passes through or very near to, the classification maximum and the maximum can be recovered from the iteration record. The legislator/chooser procedure is guaranteed to converge to a very strong local maximum. That is, a local maximum for which the point cannot be moved in any orthogonal direction and have the correct classifications increase. When the two procedures are used together in an alternating framework to analyze binary choice matrices, their performance is very good. The Monte-Carlo tests in section 4 and the empirical applications in section 5 are testimony to this fact.

References

Cahoon, Lawrence S., Melvin J. Hinich, and Peter C. Ordeshook. 1978. "A Statistical Multidimensional Scaling Method Based on the Spatial Theory of Voting." In *Graphical Representation of Multivariate Data*, edited by P. C. Wang. New York: Academic Press.

Coombs, Clyde 1964. *A Theory of Data*. New York: Wiley.

Cox, Gary and Mathew D. McCubbins. 1993. *Legislative Leviathan: Party Government in the House*. Berkeley: University of California Press.

Eckart, Carl and Gale Young. 1936. "The Approximation of One Matrix by Another of Lower Rank." *Psychometrika*, 1:211-218.

Greene, William H. 1990. *Econometric Analysis*. Englewood Cliffs, N.J.: Prentice Hall.

Heckman, James J. and James M. Snyder. 1997. "Linear Probability Models of the Demand for Attributes With an Empirical Application to Estimating the Preferences of Legislators." *Rand Journal of Economics*, 28:142-189.

King, David C. 1998. "Party Competition and Polarization in American Politics." Paper presented at the 1998 Annual Meeting of the Midwest Political Science Association, Chicago, Illinois.

Kruskal, Joseph B., Forrest W. Young, and Judith B. Seery. 1978. "How to Use KYST-2, a Very Flexible Program to do Multidimensional Scaling and Unfolding." Bell Laboratories (unpublished).

MacRae, Duncan, Jr. 1958. *Dimensions of Congressional Voting*. Berkeley: University of California Press.

Manski, Charles F. 1975. "Maximum Score Estimation of the Stochastic Utility Model of Choice." *Journal of Econometrics*, 3:205-228.

Manski, Charles F. 1985. "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator." *Journal of Econometrics*, 27:313-333.

Manski, Charles F. and T. Scott Thompson. 1986. "Operational Characteristics of Maximum Score Estimation." *Journal of Econometrics*, 32:85-108.

Poole, Keith T. and Howard Rosenthal. 1984. "U.S. Presidential Elections 1968-1980: A Spatial Analysis." *American Journal of Political Science*, 28:282-312.

Poole, Keith T. and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Rabinowitz, George. 1976. "A Procedure for Ordering Object Pairs Consistent With the Multidimensional Unfolding Model." *Psychometrika*, 45:349-373.

Ross, John and Norman Cliff. 1964. "A Generalization of the Interpoint Distance Model."" *Psychometrika*, 29:167-176.

Schonemann, Peter H. 1966. "A Generalized Solution of the Orthogonal Procrustes Problem." *Psychometrika*, 31:1-10.

Spector, L. and M. Mazzeo. 1980. "Probit Analysis and Economic Education." *Journal of Economic Education*, 11:37-44.

Wang, Ming-Mei, Peter H. Schonemann, and Jerrold G. Rusk. 1975. "A Conjugate Gradient Algorithm for the Multidimensional Analysis of Preference Data." *Multivariate Behavioral Research*, 10:45-80.

Weisberg, Herbert F. and Jerrold G. Rusk. 1970. "Dimensions of Candidate Evaluation." *American Political Science Review*, 64:1167-1185.

Young, Gale and A. S. Householder. 1938. "Discussion of a Set of Points in Terms of their Mutual Distances." *Psychometrika*, 3:19-22.

Endnotes

[1] This model was first proposed by MacRae (1958) and later developed by Poole and Rosenthal (1997) in their NOMINATE procedure.

[2] The first two dimensions estimated by NOMINATE classify about 85 percent of the roll call choices during the post World War II period (Poole and Rosenthal ,1997, ch. 2).

[3] Greene (p. 659) compared the MSE algorithm with Probit using a data set gathered by Spector and Mazzeo (1980) that studied a new method of teaching economics. Greene found that Probit correctly classified 26 of the 32 observations while the MSE algorithm only classified 22 of the 32 observations. The cutting plane procedure recovers essentially the same coefficient vector as that shown in Greene (after normalization) and correctly classifies 28 of 32 observations.

[4] Technically, given a matrix of squared distances, double-centering is subtracting from each entry in the matrix the mean of the row, the mean of the column, and adding the mean of the matrix. This has the effect of removing the squared terms from the matrix leaving just the cross-product matrix. It also reduces the rank of the matrix by one (see Young and Householder, 1938; Ross and Cliff, 1964).

Table 1A

Monte-Carlo Tests of Cutting Plane Procedure
100 Legislators and 500 Votes (Perfect Data)
(Each Entry Average of 10 Trials, Standard Deviations in Parentheses)

S	Average Majority Margin	Average Number of Errors	Average Percent Correctly Classified	Average Fit With True Normal Vectors All[a]	Average Fit With True Normal Vectors 10% Min.[b]
2	65.6 (0.7)	4.1 (3.4)	99.99 (0.01)	.998 (.001)	.999 (.000)
3	66.4 (0.5)	4.9 (2.6)	99.99 (0.01)	.996 (.000)	.997 (.000)
4	66.7 (1.0)	4.8 (1.9)	99.99 (0.00)	.992 (.001)	.995 (.001)
5	67.3 (0.7)	6.5 (2.3)	99.99 (0.00)	.989 (.001)	.993 (.000)
6	66.9 (0.7)	7.2 (1.7)	99.99 (0.00)	.984 (.001)	.990 (.000)
7	67.4 (0.7)	5.6 (2.6)	99.99 (0.01)	.979 (.002)	.987 (.001)
8	67.4 (0.7)	5.5 (2.4)	99.99 (0.01)	.975 (.001)	.984 (.001)
9	67.6 (0.5)	6.5 (2.3)	99.99 (0.00)	.970 (.002)	.980 (.001)
10	67.5 (0.8)	6.4 (3.0)	99.99 (0.01)	.964 (.001)	.976 (.001)

Table 1B

Monte-Carlo Tests of Cutting Plane Procedure
500 Votes With Normal, Uniform, and Logit Error
(Each Entry Average of 10 Trials, Standard Deviations in Parentheses)

S	P	Average Percent Error	Average Majority Margin	Average Percent Correctly Classified Obs.[c]	Average Percent Correctly Classified True[d]	Average Fit With True Normal Vectors All[e]	Average Fit With True Normal Vectors 10% Min.[f]
1	100	24.9 (0.4)	65.9 (0.6)	77.9 (0.2)	91.6 (0.5)	.840 (.035)	.894 (.015)
1	100	15.7 (2.6)	66.5 (0.8)	86.5 (0.6)	94.6 (0.3)	.906 (.026)	.939 (.007)
2	100	25.5 (0.2)	64.1 (0.2)	78.8 (0.3)	90.3 (0.3)	.951 (.004)	.952 (.003)
2	100	15.0 (0.6)	66.4 (0.8)	89.7 (0.5)	94.2 (0.2)	.979 (.004)	.986 (.001)
3	100	25.1 (0.3)	64.8 (0.6)	80.5 (0.4)	88.9 (0.3)	.913 (.008)	.918 (.006)
3	100	14.3 (0.3)	68.0 (0.5)	89.8 (0.2)	93.0 (0.2)	.954 (.004)	.969 (.002)
3	25	14.8 (0.5)	67.6 (0.6)	93.4 (0.6)	88.8 (0.4)	.890 (.011)	.909 (.013)
3	50	14.8 (0.4)	66.8 (0.7)	90.8 (0.4)	91.0 (0.3)	.934 (.008)	.952 (.003)
3	200	14.5 (0.2)	67.2 (0.8)	88.3 (0.2)	94.4 (0.2)	.970 (.002)	.980 (.002)
3	100	15.0[g] (0.4)	66.9 (0.7)	89.3 (0.4)	92.4 (0.2)	.960 (.004)	.968 (.002)
3	100	15.4[h] (0.5)	68.2 (0.5)	88.9 (0.5)	92.4 (0.3)	.952 (.003)	.965 (.003)

[a] Average cosine computed between true normal vectors and estimated normal vectors.

[a] Average cosine computed between true normal vectors and estimated normal vectors for those roll calls with at least 10 percent or better in the minority.

[c] Average correct classification of observed roll call data.

[d] Average correct classification of true roll call data.

[e] For one dimension, average Spearman correlation between estimated rank order and true rank order of midpoints. For more than one dimension, average cosine between estimated and true normal vectors.

[f] Average Spearman correlation and average cosine computed between roll calls with at least 10 percent or better in the minority.

[g] Uniform distribution error.

[h] Logit distribution error.

Table 2

Case	Ordering	Classification h	Classification a	Limits of a That Correctly Project $x_i^{(h+1)}$
1.	$-1 < c_j < w_{ij}^{(h)} < w_{ij}^{(a)} < +1$	C[1]	C	$\frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
2.	$-1 < c_j < w_{ij}^{(h)} < w_{ij}^{(a)} < +1$	I	I	$\frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
3.	$-1 < c_j < w_{ij}^{(a)} < w_{ij}^{(h)} < +1$	C	C	$\frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
4.	$-1 < c_j < w_{ij}^{(a)} < w_{ij}^{(h)} < +1$	I	I	$\frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
5.	$-1 < w_{ij}^{(h)} < w_{ij}^{(a)} < c_j < +1$	C	C	$\frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
6.	$-1 < w_{ij}^{(h)} < w_{ij}^{(a)} < c_j < +1$	I	I	$\frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
7.	$-1 < w_{ij}^{(a)} < w_{ij}^{(h)} < c_j < +1$	C	C	$\frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
8.	$-1 < w_{ij}^{(a)} < w_{ij}^{(h)} < c_j < +1$	I	I	$\frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
9.	$-1 < w_{ij}^{(h)} < c_j < w_{ij}^{(a)} < +1$	C	I	$\frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
10.	$-1 < w_{ij}^{(h)} < c_j < w_{ij}^{(a)} < +1$	I	C	$\frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
11.	$-1 < w_{ij}^{(a)} < c_j < w_{ij}^{(h)} < +1$	C	I	$\frac{1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$
12.	$-1 < w_{ij}^{(a)} < c_j < w_{ij}^{(h)} < +1$	I	C	$\frac{c_j - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}} < a_j < \frac{-1 - w_{ij}^{(h)}}{w_{ij}^{(a)} - w_{ij}^{(h)}}$

[1] "C" is correctly classified; "I" is incorrectly classified.

Table 3A
Monte-Carlo Tests of Legislator Procedure
100 Legislators and 500 Votes (Perfect Data)
(Each Entry Average of 10 Trials, Standard Deviations in Parentheses)

S	Average Majority Margin	Average Number of Errors	Average Percent Correctly Classified	Average Worst Leg. R-Square[a]	Average Best Leg. R-Square[b]
2	65.8 (1.1)	0 (0.0)	100 (0.000)	.941 (.023)	.998 (.001)
3	66.6 (0.8)	.4 (0.7)	99.999 (0.001)	.985 (.010)	.998 (.000)
4	66.5 (0.6)	1.6 (1.4)	99.997 (0.002)	.992 (.003)	.998 (.000)
5	67.4 (0.6)	3.2 (2.2)	99.99 (0.004)	.994 (.002)	.998 (.000)
6	67.2 (0.4)	5.7 (3.2)	99.99 (0.006)	.995 (.001)	.998 (.000)
7	67.9 (0.8)	13.7 (5.5)	99.98 (0.01)	.992 (.004)	.998 (.001)
8	67.4 (0.6)	21.0 (6.2)	99.96 (0.01)	.992 (.002)	.997 (.000)
9	67.4 (0.7)	30.7 (6.8)	99.94 (0.01)	.991 (.003)	.996 (.001)
10	67.4 (0.5)	42.6 (12.4)	99.91 (0.02)	.990 (.002)	.996 (.001)

Table 3B

Monte-Carlo Tests of Legislator Procedure
100 Legislators With Normal, Uniform, and Logit Error
(Each Entry Average of 10 Trials, Standard Deviations in Parentheses)

S	Q	Average Percent Error	Average Majority Margin	Average Percent Correctly Classified Obs.[c]	Average Percent Correctly Classified True[d]	Average Worst Leg. R-Square	Average Best Leg. R-Square
1	500	25.4 (0.6)	63.7 (0.6)	77.8 (0.6)	91.3 (0.4)	---	.985 (.002)
1	500	15.9 (0.3)	66.9 (0.6)	86.4 (0.4)	94.4 (0.2)	---	.985 (.012)
2	500	24.4 (0.7)	64.9 (0.3)	79.2 (0.6)	91.3 (0.2)	.967 (.016)	.984 (.006)
2	500	15.2 (0.3)	68.1 (0.7)	87.6 (0.4)	94.2 (0.2)	.971 (.013)	.991 (.003)
3	500	25.5 (0.2)	65.6 (0.5)	78.3 (0.3)	90.1 (0.3)	.943 (.016)	.972 (.008)
3	500	16.2 (0.3)	67.1 (0.6)	86.6 (0.4)	93.4 (0.2)	.968 (.012)	.985 (.003)
3	50	16.2 (0.2)	68.5 (2.1)	90.4 (0.7)	89.6 (0.9)	.725 (.052)	.819 (.018)
3	100	16.1 (0.4)	67.8 (1.1)	88.6 (0.3)	91.2 (0.4)	.835 (.016)	.888 (.018)
3	250	16.1 (0.4)	66.9 (0.8)	87.1 (0.3)	92.7 (0.3)	.930 (.016)	.955 (.008)
3	500	14.8[e] (0.5)	67.1 (0.7)	88.0 (0.4)	93.5 (0.2)	.968 (.009)	.982 (.007)
3	500	15.3[f] (0.2)	68.1 (0.9)	87.4 (0.3)	93.4 (0.2)	.968 (.011)	.986 (.004)

[a] R-Squares computed between true and reproduced legislator coordinates. The number shown is the average of the worst r-squares across the 10 trials.

[b] R-Squares computed between true and reproduced legislator coordinates. The number shown is the average of the best r-squares across the 10 trials.

[c] Average correct classification of observed roll call data.

[d] Average correct classification of true roll call data.

[e] Uniform distribution error.

[f] Logit distribution error.

Table 4

Monte-Carlo Tests: Non-Parametric Unfolding of Binary Choice Matrices
100 Legislators and 500 Votes
(Each Entry Average of 10 Trials, Standard Deviations in Parentheses)

S	Average Majority Margin	Average Number of Errors	Average Percent Correctly Classified	Average Worst Leg. R-Square	Average Best Leg. R-Square[a]	Average Fit With True Normal Vectors All	Average Fit With True Normal Vectors 10% Min.
1	68.2 (1.3)	0 (0.0)	100.00 (0.00)	---	1.000 (.000)	1.000 (.000)	1.000 (.000)
2	65.2 (0.8)	22.5 (10.2	99.96 (0.02)	.940 (.014)	.979 (.005)	.995 (.002)	.996 (.002)
3	66.0 (1.0)	10.1 (4.0)	99.98 (0.01)	.964 (.016)	.981 (.007)	.991 (.002)	.993 (.002)
4	66.7 (1.2)	6.7 (3.0)	99.99 (0.01)	.967 (.011)	.983 (.004)	.987 (.003)	.990 (.003)
5	66.8 (0.6)	6.5 (3.0)	99.99 (0.01)	.970 (.005)	.980 (.003)	.984 (.001)	.987 (.001)
6	67.0 (0.7)	5.3 (2.5)	99.99 (0.00)	.959 (.006)	.978 (.003)	.978 (.001)	.987 (.001)
7	67.3 (0.6)	6.5 (3.5)	99.99 (0.01)	.961 (.004)	.977 (.003)	.972 (.002)	.979 (.001)
8	67.2 (0.5)	7.5 (2.2)	99.99 (0.00)	.953 (.012)	.976 (.003)	.966 (.002)	.975 (.002)
9	67.8 (0.7)	6.6 (1.3)	99.99 (0.00)	.952 (.008)	.971 (.003)	.960 (.002)	.970 (.003)
10	67.8 (0.8)	6.7 (1.8)	99.99 (0.00)	.943 (.012)	.972 (.003)	.957 (.002)	.966 (.002)

[a] For s=1, the squared Spearman Rank Correlation is computed between the 100 true and reproduced legislator ranks.

Table 5
Monte-Carlo Tests: Non-Parametric Unfolding of Binary Choice Matrices With Missing Data
(Each Entry Average of 10 Trials, Standard Deviations in Parentheses)
2 Dimensions, 100 Legislators, 500 Votes

Percent Missing	Average Percent Error	Average Majority Margin	Percent Correct 1 Dim.	Percent Correct 2 Dim.	Percent Correct 3 Dim.	Percent Correct 4 Dim	Percent Correct 5 Dim.	R^2 1st	R^2 2nd
0	0	65.5	91.4	99.9	100.0	100.0	100.0	.982	.948
		(0.7)	(0.9)	(0.0)	(0.0)	(0.0)	(0.0)	(.005)	(.005)
20	0	65.5	91.6	99.9	100.0	100.0	100.0	.983	.949
		(0.7)	(1.5)	(0.0)	(0.0)	(0.0)	(0.0)	(.005)	(.009)
50	0	65.9	91.6	99.9	100.0	100.0	100.0	.975	.938
		(0.8)	(0.9)	(0.0)	(0.0)	(0.0)	(0.0)	(.005)	(.020)
70	0	66.5	92.1	99.9	100.0	100.0	100.0	.964	.920
		(0.8)	(0.7)	(0.0)	(0.0)	(0.0)	(0.0)	(.007)	(.013)
0	20.3	64.6	79.3	84.0	85.1	86.0	86.6	.968	.953
	(0.6)	(0.8)	(0.7)	(0.4)	(0.3)	(0.3)	(0.3)	(.007)	(.004)
20	20.3	64.8	79.3	84.7	86.0	86.8	87.5	.963	.936
	(0.6)	(0.8)	(0.8)	(0.4)	(0.3)	(0.3)	(0.3)	(.005)	(.012)
50	20.3	65.3	80.5	85.9	87.5	88.7	89.5	.949	.925
	(0.6)	(0.8)	(0.5)	(0.4)	(0.3)	(0.4)	(0.4)	(.012)	(.012)
70	20.3	66.1	81.1	87.6	89.7	91.3	92.8	.929	.915
	(0.6)	(0.9)	(0.8)	(0.5)	(0.5)	(0.5)	(0.5)	(.009)	(.012)

3 Dimensions, 100 Legislators, 500 Votes

Percent Missing	Average Percent Error	Average Majority Margin	Percent Correct 1 Dim.	Percent Correct 2 Dim.	Percent Correct 3 Dim.	Percent Correct 4 Dim	Percent Correct 5 Dim.	R^2 1st	R^2 2nd	R^2 3rd
0	0	66.7	85.3	92.7	100.0	100.0	100.0	.985	.976	.957
		(0.7)	(0.4)	(0.8)	(0.0)	(0.0)	(0.0)	(.005)	(.009)	(.010)
20	0	66.6	85.5	92.3	100.0	100.0	100.0	.981	.976	.958
		(0.8)	(0.5)	(0.9)	(0.0)	(0.0)	(0.0)	(.002)	(.004)	(.013)
50	0	67.0	85.6	92.9	99.9	100.0	100.0	.962	.946	.924
		(0.7)	(0.5)	(0.9)	(0.0)	(0.0)	(0.0)	(.009)	(.013)	(.015)
70	0	68.0	86.3	93.4	99.9	100.0	100.0	.932	.911	.883
		(0.4)	(0.3)	(0.8)	(0.0)	(0.0)	(0.0)	(.012)	(.019)	(.022)

0	22.8	61.3	73.8	77.7	81.2	81.9	82.6	.982	.980	.979
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(.000)	(.000)	(.000)
20	22.8	61.7	74.1	78.3	81.9	82.6	83.3	.981	.977	.974
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(.000)	(.000)	(.000)
50	22.8	62.4	74.9	79.9	83.4	84.3	85.3	.969	.966	.952
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(.000)	(.000)	(.000)
70	22.8	63.4	76.4	82.1	85.6	87.1	88.5	.947	.938	.908
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(.000)	(.000)	(.000)

Table 6
U.S. Senate: 1947 - 1996
Non-Parametric Unfolding of Roll Call Data

Senate	Years	Senators	Roll Calls	Total Choices	Average Margin	Non-P One	Non-P Two	kyst R 1st	kyst R 2nd	nom R 1st	nom R 2nd
104	1995-96	103[a]	805[b]	78,882[c]	.637	90.0[d]	91.3	.985[e]	.656	.980[f]	.785
103	1993-94	101	647	63,023	.672	89.2	90.4	.984	.744	.983	.818
102	1991-92	102	481	46,208	.685	86.9	88.5	.983	.812	.979	.879
101	1989-90	101	499	48,649	.680	85.4	87.1	.990	.860	.985	.884
100	1987-88	101	635	59,631	.709	87.7	89.5	.981	.722	.981	.925
99	1985-86	101	661	63,104	.688	84.7	86.8	.994	.841	.976	.936
98	1983-84	101	578	53,330	.698	84.8	87.3	.990	.900	.979	.954
97	1981-82	101	818	77,672	.682	85.5	88.1	.995	.897	.987	.956
96	1979-80	101	928	82,937	.683	83.5	85.8	.988	.803	.988	.965
95	1977-78	104	1037	92,868	.691	84.5	86.4	.989	.757	.977	.864
94	1975-76	100	1144	100,328	.691	86.3	88.6	.990	.888	.982	.934
93	1973-74	101	983	87,699	.695	85.1	87.5	.993	.908	.991	.961
92	1971-72	102	783	68,588	.676	85.0	88.6	.991	.944	.981	.971
91	1969-70	102	557	49,219	.681	84.5	88.1	.991	.898	.984	.973
90	1967-68	101	518	46,081	.699	83.6	87.2	.988	.899	.988	.949
89	1965-66	102	441	40.618	.681	85.4	88.4	.988	.901	.975	.912
88	1963-64	102	505	47,797	.686	85.0	90.1	.974	.963	.913	.937
87	1961-62	105	400	38,189	.675	87.3	90.6	.947	.960	.963	.933
86	1959-60	103	360	33,855	.686	84.9	89.6	.976	.962	.963	.956
85	1957-58	98	255	23,097	.669	84.7	89.5	.982	.924	.974	.895
84	1955-56	99	184	16,798	.659	85.5	90.4	.980	.927	.975	.925
83	1953-54	103	242	20,991	.672	86.9	90.3	.949	.731	.950	.906
82	1951-52	96	208	17,368	.659	86.0	89.4	.961	.769	.977	.928
81	1949-50	102	447	38,074	.667	85.0	88.6	.969	.862	.957	.955
80	1947-48	97	237	20,321	.665	88.0	90.8	.970	.885	.961	.917

[a] Number of Senators may exceed two times the number of States because of within Congress replacements.
[b] Number of roll calls with at least 2.5% voting, paired, or announced, on losing side.
[c] Total choices may not equal number of Senators times number of roll calls because of non-voting due to absences, etc..
[d] Classifications from non-parametric unfolding algorithm.
[e] Squared Pearson correlation between Senator coordinates from KYST and Senator coordinates from non-parametric unfolding. Non-parametric unfolding configuration rotated to best match KYST configuration.
[f] Squared Pearson correlation between Senator coordinates from W-NOMINATE and Senator coordinates from non-parametric unfolding. Non-parametric unfolding configuration rotated to best match W-NOMINATE configuration.

Table 7

104th (1995-96) U.S. Senate

Name	Rank	Name	Rank	Name	Rank
Simon (D-IL)	1	Baucus (D-MT)	45	Dole (R-KS)	89
Wellstone (D-MN)	2	Nunn (D-GA)	46	Lott (R-MS)	90
Feingold (D-WI)	3	Heflin (D-AL)	47	Craig (R-ID)	91
Wyden (D-OR)	4	Campbell (D-CO)	48	Kempthorne (R-ID)	92
Boxer (D-CA)	5	Jeffords (R-VT)	49	Grams (R-MN)	93
Kennedy (D-MA)	6	Cohen (R-ME)	50	Nickles (R-OK)	94
Moseley-Braun (D-IL)	7	Specter (R-PA)	51	Smith (R-NH)	95
Levin (D-MI)	8	Snowe (R-ME)	52	McCain (R-AZ)	96
Bradley (D-NJ)	9	Chafee (R-RI)	53	Ashcroft (R-MO)	97
Lautenberg (D-NJ)	10	Hatfield (R-OR)	54	Inhofe (R-OK)	98
Leahy (D-VT)	11	Campbell (R-CO)	55	Gramm (R-TX)	99
Murray (D-WA)	12	Kassebaum (R-KS)	56	Helms (R-NC)	100
Bumpers (D-AR)	13.5	Packwood (R-OR)	57	Faircloth (R-NC)	101
Harkin (D-IA)	13.5	Simpson (R-WY)	58	Brown (R-CO)	102
Pell (D-RI)	15	Roth (R-DE)	59	Kyl (R-AZ)	103
Kerry (D-MA)	16	Stevens (R-AK)	60		
Pryor (D-AR)	17	Gorton (R-WA)	61		
Moynihan (D-NY)	18	D'Amato (R-NY)	62		
Kohl (D-WI)	19	Dewine (R-OH)	63		
Sarbanes (D-MD)	20	Domenici (R-NM)	64		
Rockefeller (D-WV)	21	Lugar (R-UT)	65		
Akaka (D-HI)	22	Bond (R-MO)	66		
Daschle (D-SD)	23	Cochran (R-MS)	67		
Dodd (D-CT)	24	Murkowski (R-AK)	68		
Mikulski (D-MD)	25	Warner (R-VA)	69		
Glenn (D-OH)	26	Pressler (R-SD)	70		
Conrad (D-ND)	27	Bennett (R-UT)	71		
Dorgan (D-ND)	28	Frist (R-TN)	72		
Bingaman (D-NM)	29	Santorum (R-PA)	73		
Biden (D-DE)	30	Shelby (R-AL)	74		
Byrd (D-WV)	31	Burns (R-MT)	75		
Kerrey (D-NE)	32	Hatch (R-UT)	76		
Bryan (D-NV)	33	Abraham (R-MI)	77		
Graham (D-FL)	34	Gregg (R-NH)	78		
Feinstein (D-CA)	35	Thompson (R-TN)	79		
Hollings (D-SC)	36	Grassley (R-IA)	80		
Inouye (D-HI)	37	Thurmond (R-SC)	81		
Ford (D-KY)	38	McConnell (R-KY)	82		
Reid (D-NV)	39	Hutchison (R-TX)	83.5		
Robb (D-VA)	40	Thomas (R-WY)	83.5		
Exon (D-NE)	41	Frahm (R-KS)	85		
Lieberman (D-CT)	42	Mack (R-FL)	87		
Breaux (D-LA)	43	Coverdell (R-GA)	87		
Johnston (D-LA)	44	Coats (R-IN)	87		

Table 8
Distribution of Bootstrap Standard Errors for Senators, 85th Senate

Range of Bootstrap Standard Errors	Number of Senators First Dimension	Number of Senators Second Dimension
.00 - .05	13	0
.05 - .10	67	36
.10 - .15	14	36
.15 - .20	2	17
.20 - .25	2	4
.25 - .30	0	5

Figure 1A. Ten Point Example
Original Positions in Two Dimensions

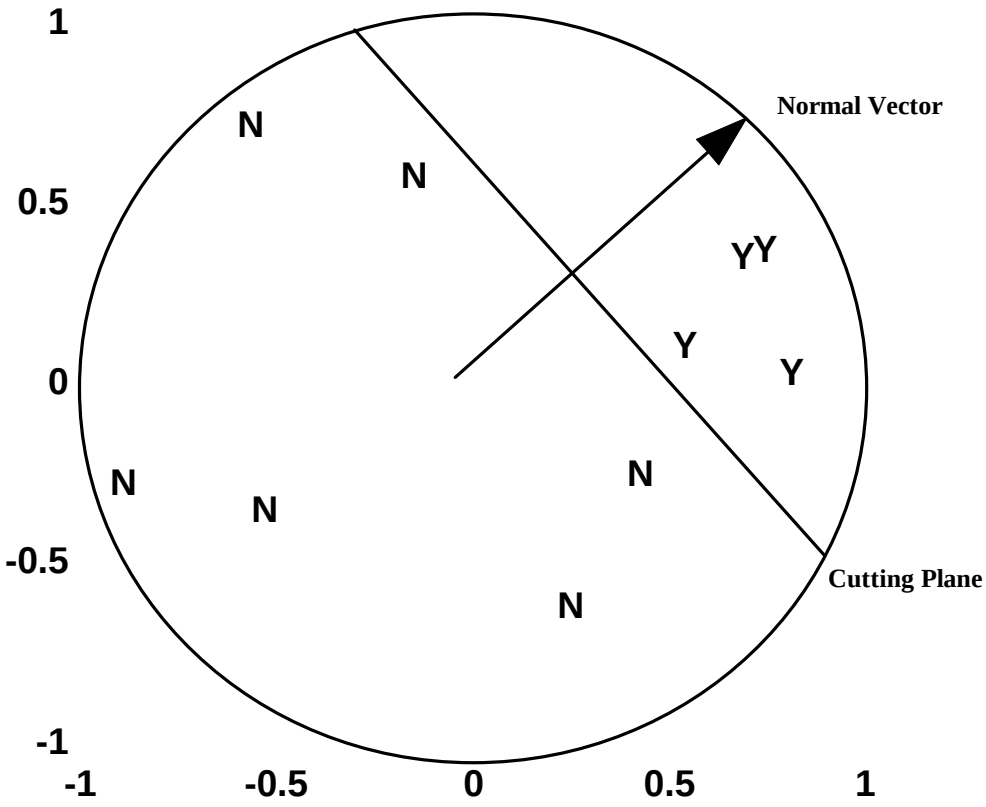


Figure 1B. Ten Point Example

Points Projected Onto Line

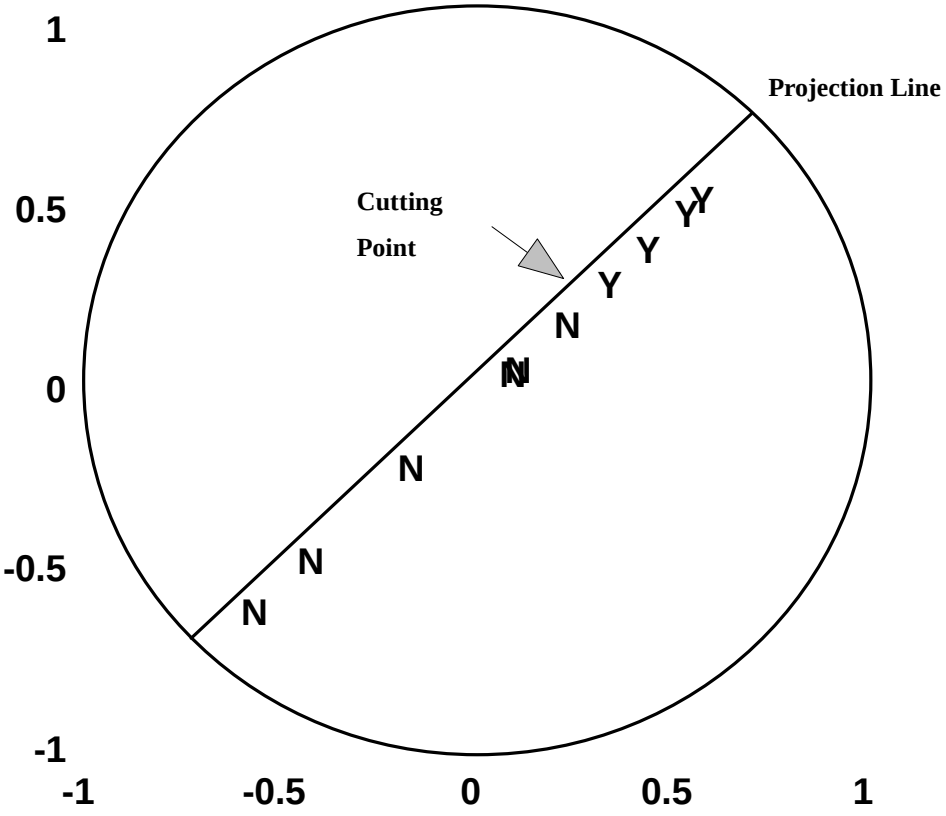


Figure 2. Calculating Correct Classification

Actual Voting Pattern

Y Y Y Y Y Y . . . . Y Y * N Y * N Y * N N . . .N N

-1.0 w_1 w_2 w_3 w_4 . . . . 0.0 . . . . w_{p-1} w_p +1.0

Perfect Voting Patterns

(-1 , w_1) produces	**nnnnnnnnn.....nn**	**or**	**yyyyyyyyy.....yy**
(w_1 , w_2) produces	**ynnnnnnnn.....nn**	**or**	**nyyyyyyyy.....yy**
(w_2 , w_3) produces	**yynnnnnnn.....nn**	**or**	**nnyyyyyyy.....yy**
(w_3 , w_4) produces	**yyynnnnnn.....nn**	**or**	**nnnyyyyyy.....yy**

etc.

(w_{p-1} , w_p) produces	**yyyyyyyyy.....yn**	**or**	**nnnnnnnnn.....ny**
(w_p , +1) produces	**yyyyyyyyy.....yy**	**or**	**nnnnnnnnn.....nn**

Figure 3A. Cutting Plane Example

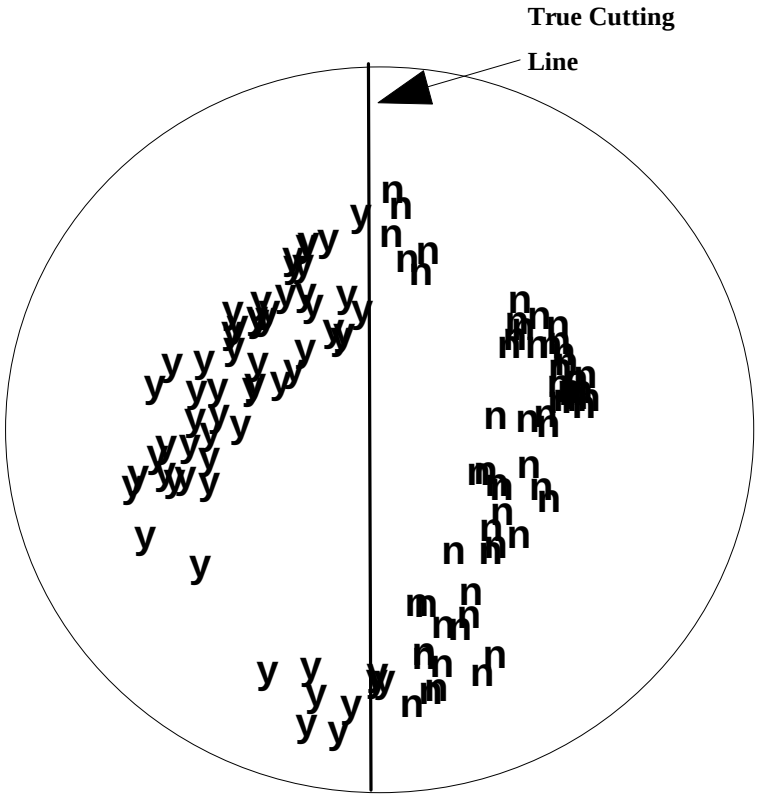


Figure 3B. Cutting Plane Example

Initial Estimate of Cutting Plane

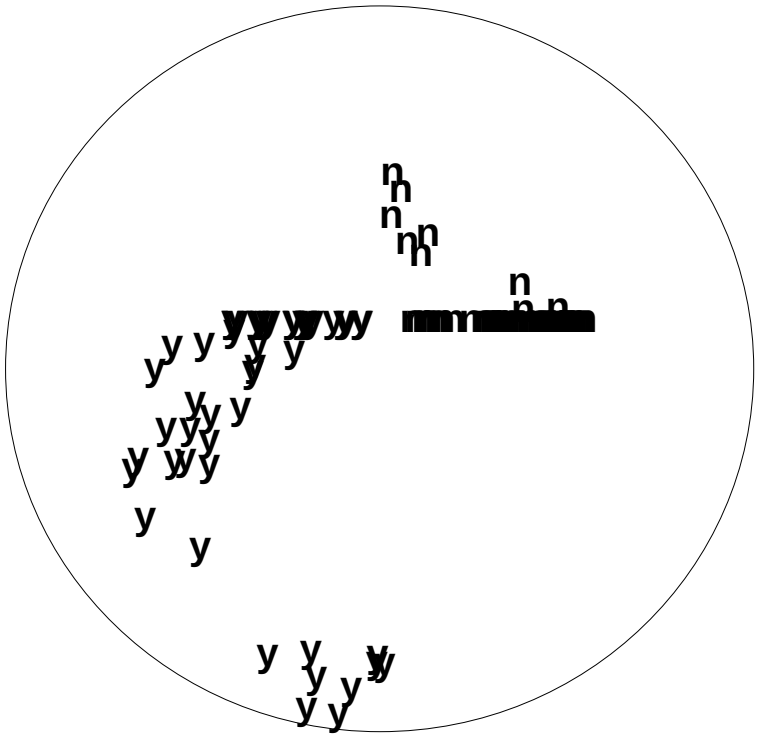


Figure 3C. Cutting Plane Example
10th Estimate of Cutting Plane

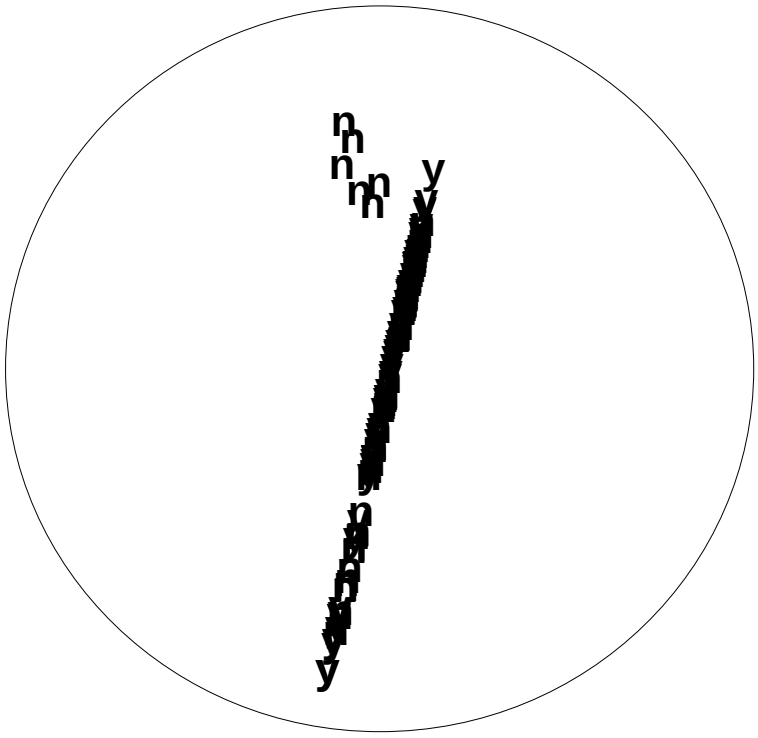


Figure 4A. Error Example

78 Errors With True Cutting Line

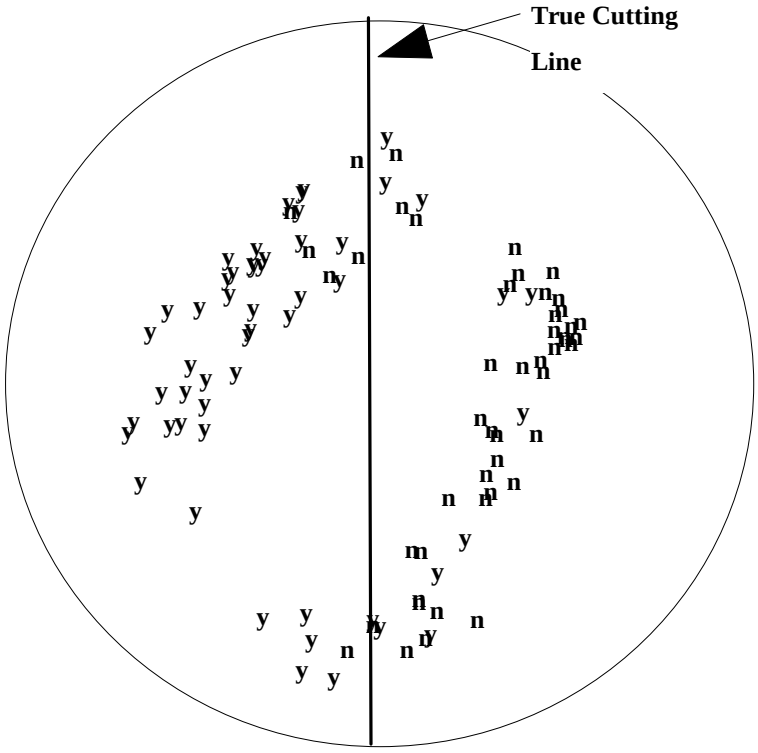


Figure 4B. Error Example
73 Errors at 10th Estimate (Minimum)

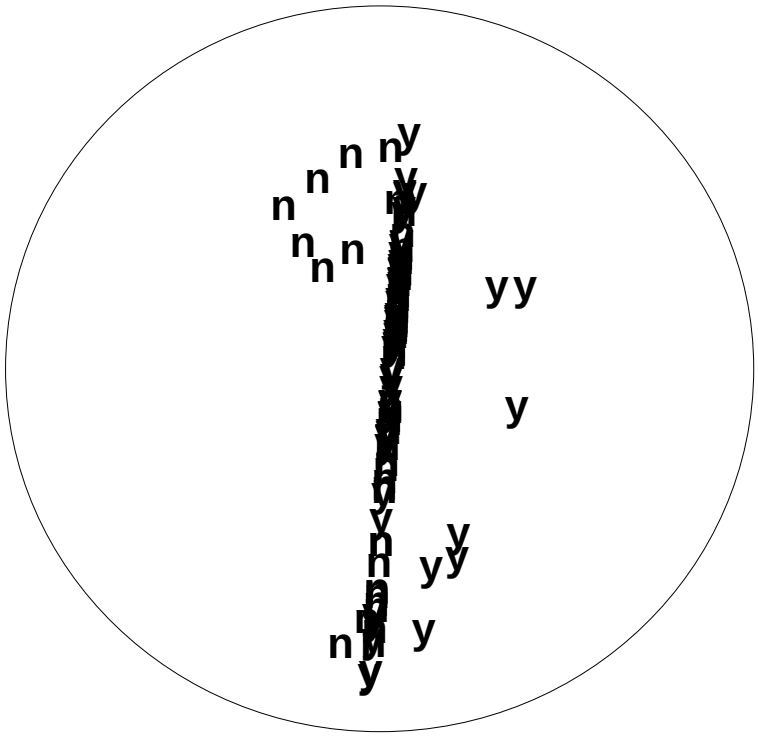


Figure 4C. Error Example
73 Errors at 20th Estimate (Minimum)

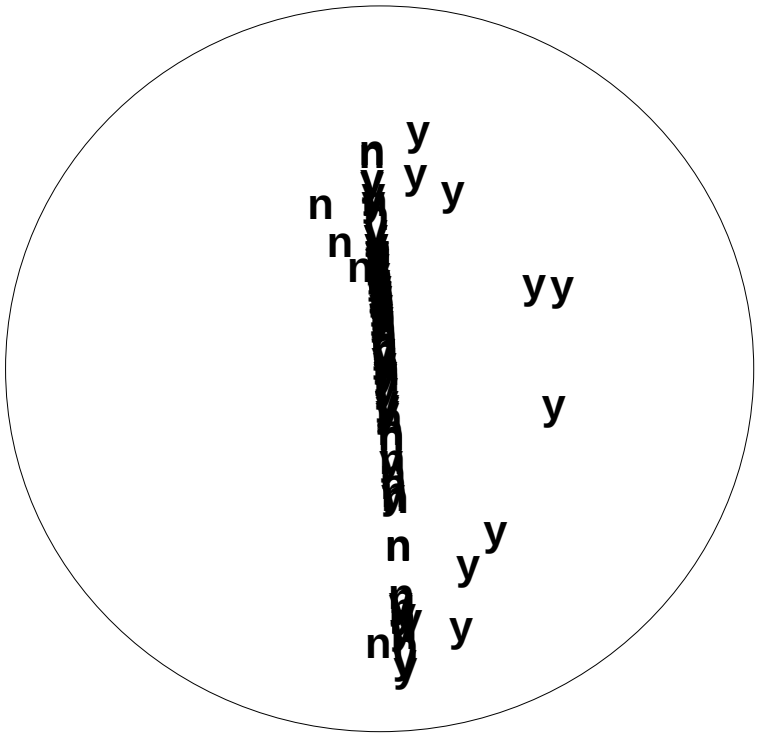


Figure 4D. Error Example

75 Errors at 30th Estimate (Converged)

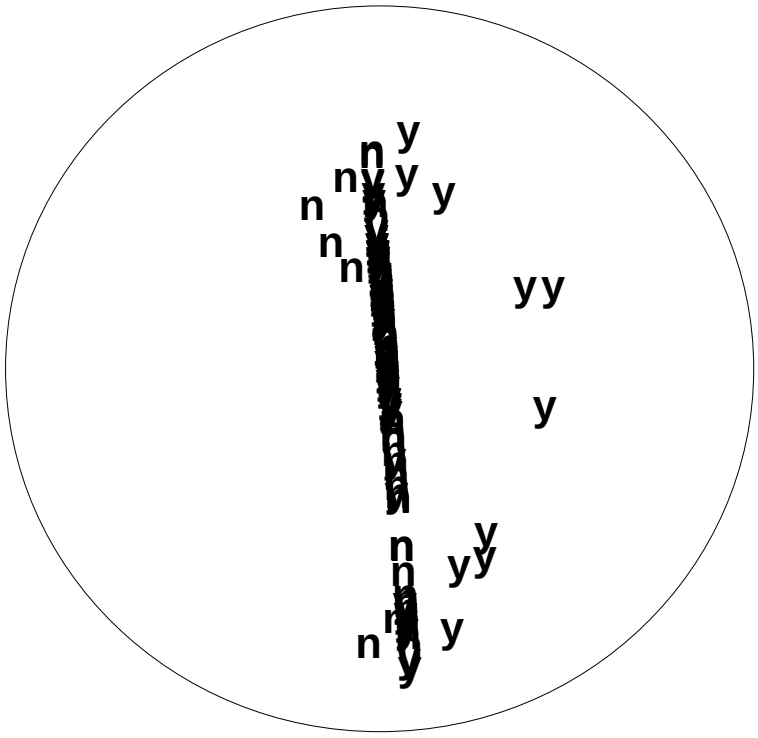


Figure 5. Legislator Example: nnnyn

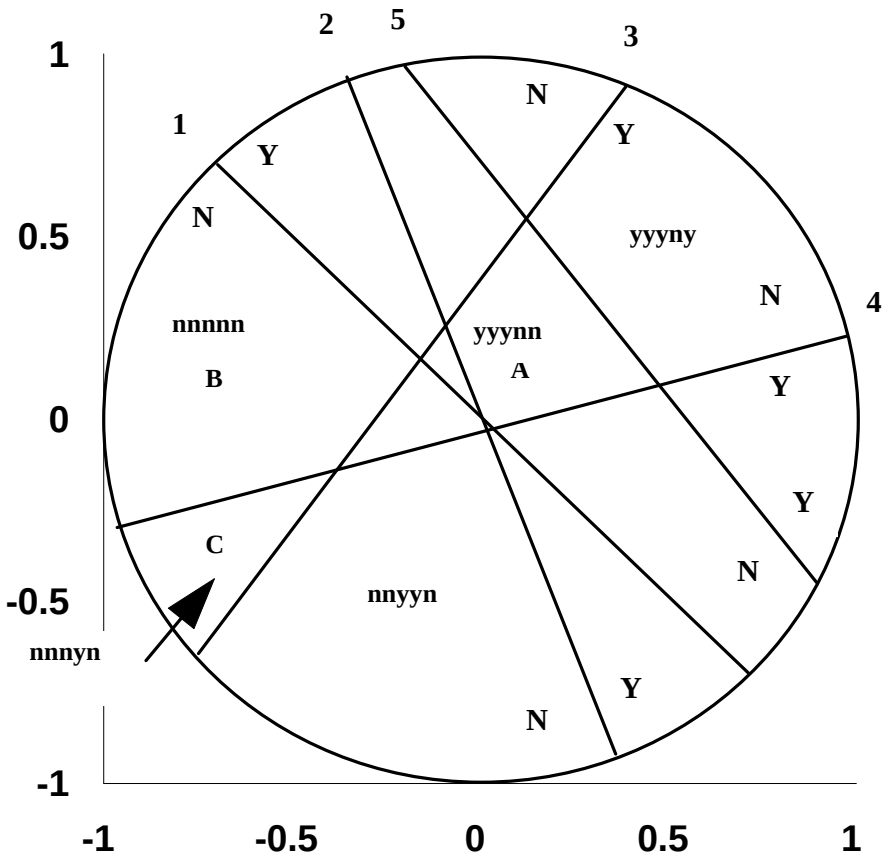
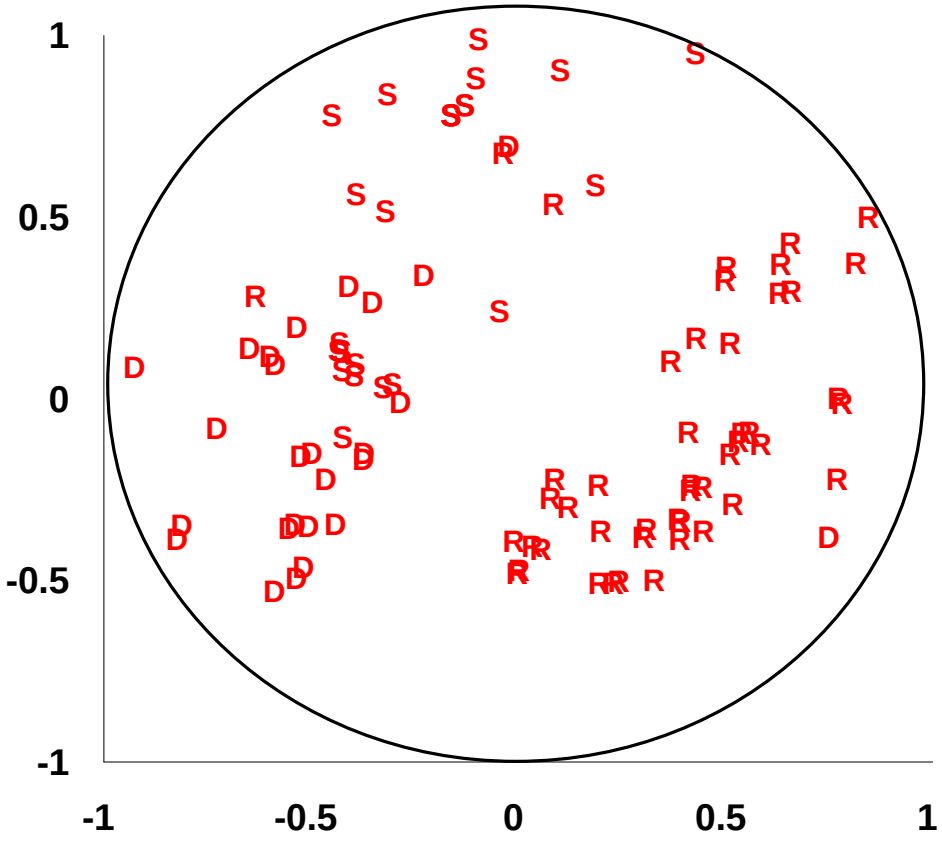


Figure 6A. 85th Senate 1957-58

Senator Locations



D=No. Democrat, S=So. Democrat

R=Republican

Figure 6B. 85th Senate 1957-58

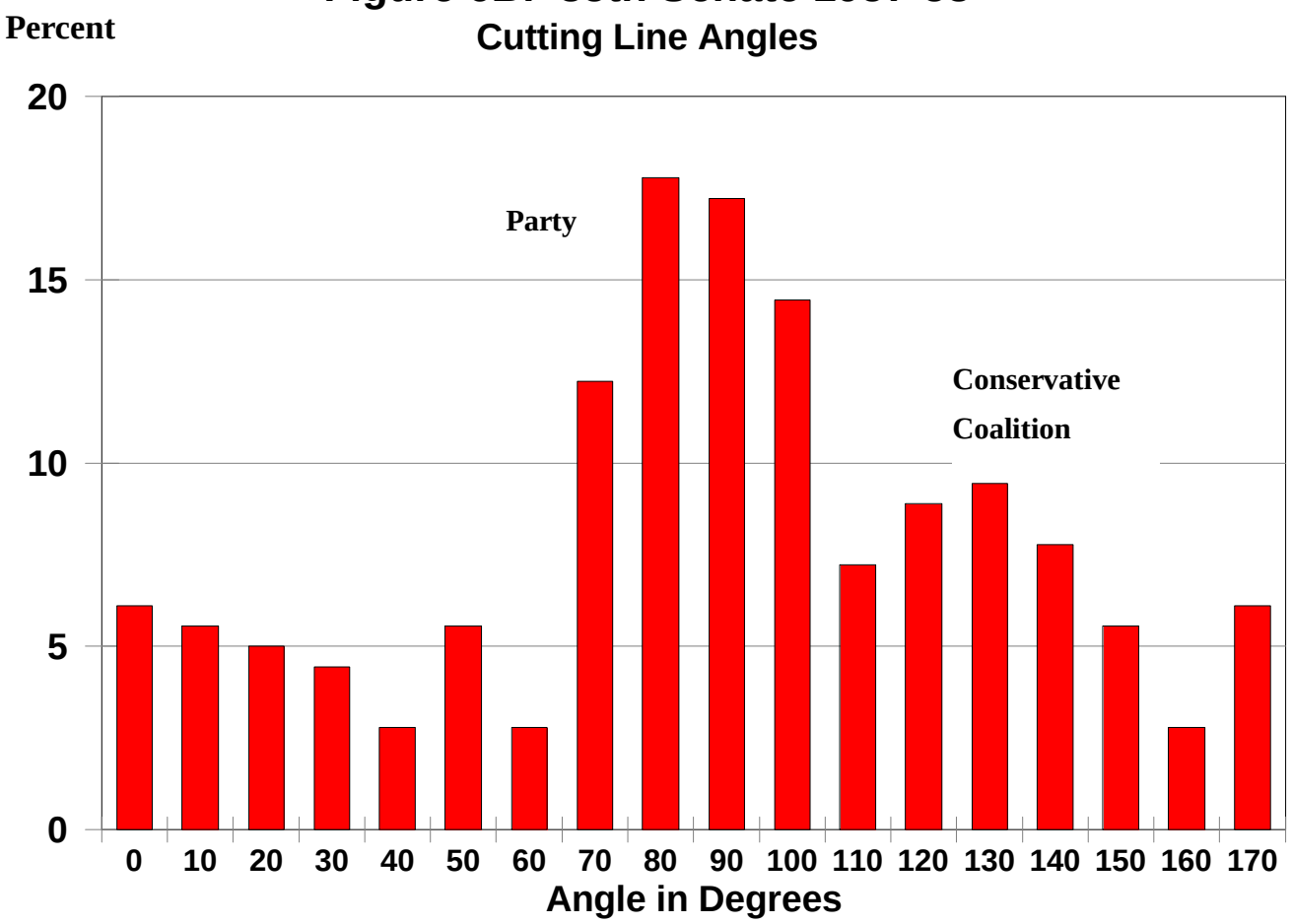


Figure 7. 1968 Feeling Thermometer Example