Mathl. Comput. Modelling Vol. 16, No. 8/9, pp. 85–101, 1992
Printed in Great Britain.

0895-7177/92 $5.00 + 0.00

EVALUATING DIMENSIONALITY IN SPATIAL VOTING MODELS

KEITH T. POOLE, FALLAW B. SOWELL, STEPHEN E. SPEAR
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, PA 15213, U.S.A.

Abstract—Spatial models of voting are widely used in Economics and Political Science. Spatial theory is an attractive framework to analyze choice because Euclidean geometry is easy to visualize and the language of politics is full of spatial references. Empirical tests have generally supported the theory but the estimation methods employed to produce the spatial representations of voters have raised serious statistical issues which have not been fully resolved. One of these issues is determining the number of dimensions. This is a difficult problem because the number of estimated parameters increases with the number of dimensions. We show a solution for this problem in this paper. We assume a uniform distribution of voters through an N-dimensional unit hypersphere with perfect spatial voting and equally salient dimensions. We then solve for the projection of this perfect voting onto one dimension and the resultant classification error. We derive the probability density function of the classification errors which can be used to calculate the likelihood that a sample of votes was drawn from an N-dimensional hypersphere. Our Monte-Carlo investigation into the properties of this likelihood function shows that it can be used reliably for low N with small sample sizes.

1. INTRODUCTION

Spatial voting models are widely used in Economics and Political Science. The early work of Anthony Downs [1957] and Duncan Black [1958] sparked the development of a large body of theory which uses simple geometric representations of individual preferences to model policy choice in legislatures and in the mass electorate.[1] Spatial theory is an attractive framework to analyze choice because Euclidean geometry is easy to visualize and the language of politics is replete with spatial references.[2] Empirical work has generally supported the theory.[3]

The dimensions of a spatial map of individual preferences represent the fundamental attributes that individuals use to evaluate policy alternatives. In our view, the best way to think about what these dimensions mean is to place them in the context of the belief-systems model of Converse [1964]. Converse defines a belief system "as a configuration of ideas and attitudes in which the elements are bound together by some form of constraint or functional interdependence" [1964, p. 207]. In the real world of politics, constraint means that certain issue positions are bundled together, and the knowledge of one or two issue positions makes the remaining positions very predictable. For example, to know that a member of the House voted against the Persian Gulf War and in favor of the 1991 Civil Rights law, makes it highly likely that the member is in favor of increased spending to aid the homeless and opposed aid to the Nicaraguan Contras during the 1980s. These relationships between issues are neatly summarized by the words "liberal" and

[1] See Krehbiel [1988] for an excellent bibliographic essay on the spatial theory of legislatures. For a general essay on spatial theory, see Austen-Smith [1983]. Ordeshook [1986] puts spatial theory within the context of game theory and integrates it with political theory in general. For more recent developments, see Calvert [1985], Banks [1990], and [Austen-Smith and Banks 1989, 1990].

[2] Consider the terms: "left," "right," "moving left," "centrist," and so on.

[3] See the extensive analyses of Enelow and Hinich [1984] on U.S. elections. Earlier work on U.S. elections was done by Weisberg and Rusk [1970], Rusk and Weisberg [1972], Rabinowitz [1976], Cahoon [1978], and Poole and Rosenthal [1984]. Applications to congressional voting include MacRae [1958, 1970], Krehbiel and Rivers [1989, 1990], and Poole and Rosenthal [1989, 1991].

"conservative" and followers of U.S. politics can easily list the issue positions normally associated with these words.

It is important to note that "liberal" and "conservative" do not necessarily indicate a *coherent* political philosophy. As Hinich and Pollard [1981, p. 325] point out, it is not necessary that this observed consistency derive "from anything so strong as ideology ... The behavioral labels we observe are more likely derived from *ad hoc* responses to changing conditions and political fortunes."

In any event, the purpose of our paper is not to delve into the nature of "ideology." There is now a fairly large body of literature which finds low dimensionality in various types of preferential data. The question we address below is how to determine the number of these dimensions. In terms of the theory of individual choice, the number of dimensions is irrelevant to the mathematical apparatus of the model. However, determining the number of dimensions in empirical applications is quite another matter.

Given a matrix of individual preferences or choices, a variety of multivariate techniques—factor analysis, multi-dimensional scaling, cluster analysis—can be used to estimate dimensions. Usually, one, two, and three (or more) dimensional models are estimated and a measure of fit (r-square, STRESS, the number of eigenvalues greater than one, etc.) is used to determine the "best" compromise between parsimony and explanatory power. Because increasing the number of dimensions always increases the fit of these techniques, the decision of the researcher that the dimensionality is k rather than $k-1$ or $k+1$ is, in effect, a decision that the increase in fit from $k-1$ to k dimensions is "significant" while the increase from k to $k+1$ dimensions is not "significant."

In this regard, estimating the number of dimensions is like simple OLS in that a perfect fit can be obtained by having as many independent variables (dimensions) as there are observations (voters). However, in simple OLS whenever an independent variable is added to the model only one additional parameter must be estimated and standard statistical theory can be used to determine the "significance" of the contribution of the additional variable. In contrast, estimating a spatial voting model requires that ideal points for every voter be estimated for every dimension. Consequently, adding a dimension represents an increase in the number of parameters equal to the number of voters. This presents some obvious statistical problems for determining the "significance" of the added dimension.[4]

In addition, simple OLS rests on the presumption that the model is correctly specified so that statements about the fit of the model and the significance of the estimated parameters are made against this "fixed" background. In contrast, spatial voting models specify a certain sort of choice behavior given a well defined geometry—the number of dimensions is not specified.

Given these thorny problems, determining dimensionality in spatial voting models empirically is an unsettled and controversial topic [Poole, 1988; Koford, 1989; Poole and Rosenthal, 1990]. The purpose of this paper is to show a new way of approaching the problem of determining dimensionality within a maximum likelihood framework.

Our work is motivated in part by Koford's [1989] observation that a useful null model for evaluating estimated spatial models is to assume that the true space is of high dimensionality with perfect spatial voting. The projection of this perfect voting down onto a lower dimensional space equal to the number of dimensions being estimated provides a useful comparative benchmark. In particular, the approach we take is to assume a uniform distribution of voters within an n-dimensional hypersphere of radius one. We assume perfect spatial voting and solve for the projection of the perfect votes onto a single dimension and the resultant classification error. The probability density function we derive can be used to compute the likelihood—based upon the marginals of the votes and the observed classification percentages—that a sample of votes was *drawn from an* n-dimensional hypersphere.

[4] A tempting, but illusory, fix for this problem is not to estimate individual ideal points at all but to estimate the parameters of a *distribution* from which the ideal points are assumed to be drawn. This solves the parameter proliferation problem in that, like simple OLS, only one (or two) additional parameters are estimated for each dimension. However, this is a case of throwing the baby out with the bathwater. The whole point of using spatial models is to learn something about choice behavior of individuals. This information is simply thrown away by taking this approach.

In Section 2 we sketch in some detail how our model works in two dimensions and then turn to a discussion of the numerical results for any value of N. (The detailed derivations for the N-dimensional case are shown in the Appendix.) In Section 3, we show a Monte Carlo study of the likelihood function and, in Section 4, we show an empirical application to congressional voting.

2. THE PROJECTION OF PERFECT N-DIMENSIONAL VOTING ONTO ONE DIMENSION

A. The Two-Dimensional Case

Assume that the voters are uniformly distributed over the unit circle and that all votes are binary—the voters are always faced with only two alternatives on any particular vote. In addition, assume that the alternatives are drawn from a uniform distribution over the unit circle and that voters vote for the closest alternative in terms of Euclidean distance. As an example, suppose that all votes are ties—50 percent to 50 percent. Figure 1 shows how several such votes would project onto a single dimension.

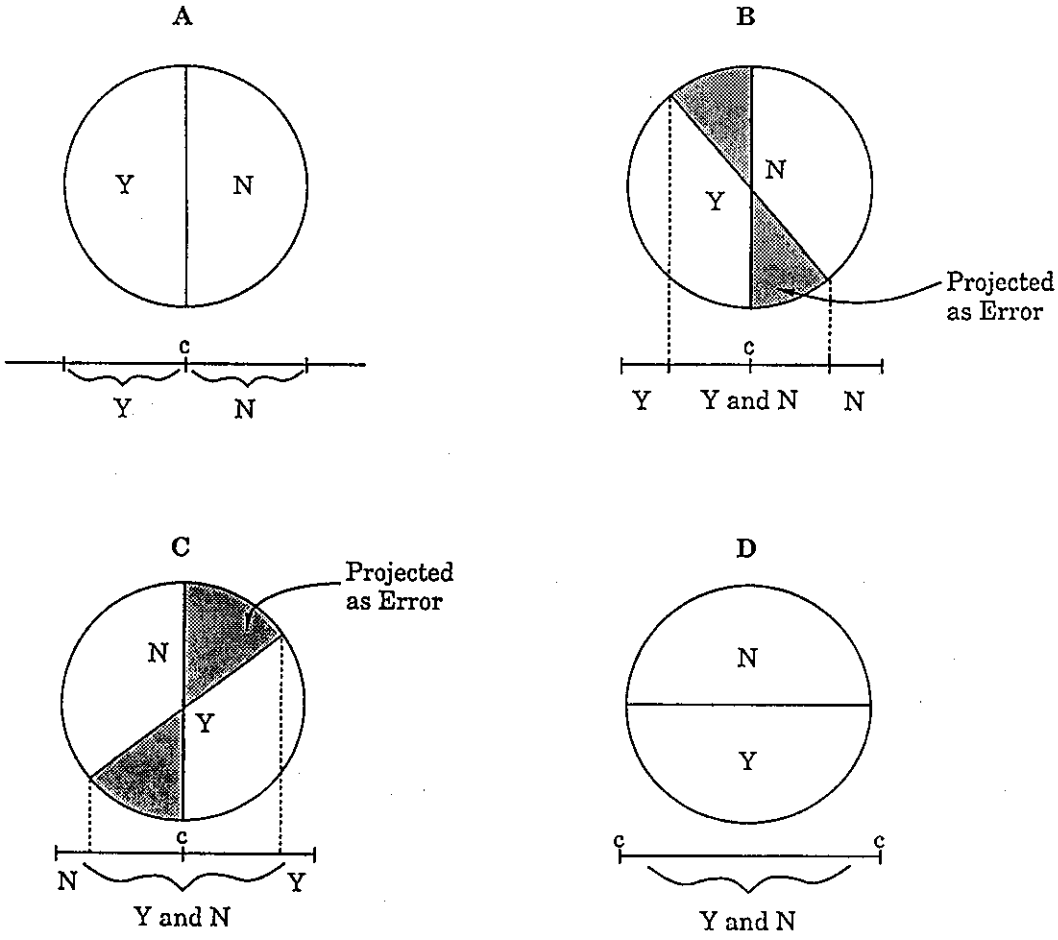

Figure 1. Projection of votes.

The line through each circle in Figure 1 is the perpendicular bisector of the line joining the two policy alternatives. Note that the angle of this "cutting line" determines how noisy the projection is onto one dimension. In Figure 1A the cutting line is perpendicular to the projection dimension so that the cutting point, c, perfectly divides the Y's from the N's. The percentage of the individual votes correctly classified would be 100 percent. In Figures 1B and 1C, the angle of the cutting line is 45 and −45 degrees, respectively. These two projections produce the same cutting point as in Figure 1A but the percentage correctly classified falls to 75 percent. To see

this, note that the shaded regions are areas of the plane which, when projected onto the single dimension, will be incorrectly classified. These shaded "pie slices" make up 25 percent of the area of the circles. Finally, in Figure 1D the cutting line is parallel to the projection dimension so that only 50 percent—which is simply equal to the marginals of the vote—is "correctly" projected. The cutting point can be at either end of the projection dimension.

The general case for a two-dimensional projection is shown in Figure 2. Without loss of generality, we can set the origin of the space to the center of the circle and place the projection dimension through the origin. The shaded portion of the circle represents the proportion of the voters in the minority. The angle ϕ and the distance from the origin, r_m, determine the orientation of the minority region to the projection dimension and the size of the minority region respectively. The cutting point is given by $r_m \sec \phi$ and the "pie slices" produced by the line through the cutting point and perpendicular to the projection dimension constitute the classification error when projected down onto the dimension. The cutting point minimizes the classification error. This fact is easily shown through a simple exercise in geometry (see the Appendix).

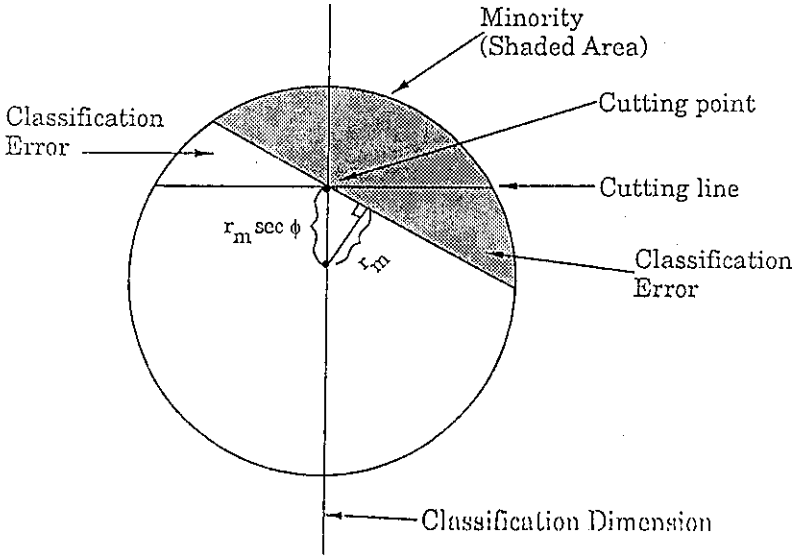

Figure 2. General case of two-dimensional projection.

In terms of Figure 2, the minority proportion, m, is a function solely of the distance of the cutting line from the origin; namely

$$m = \frac{\cos^{-1} r_m - r_m\sqrt{1 - r_m^2}}{\pi}. \tag{1}$$

In contrast, the classification error—the ratio of the sum of the two "pie slices" to the area of the unit circle—is a function of r_m and ϕ. In terms of ϕ, two cases have to be considered. First, as shown in Figure 2, ϕ is such that the cutting line intersects the projection dimension; and second, ϕ is large and the cutting line does not intersect the projection dimension. When ϕ is such that the cutting line exactly intersects the end of the projection dimension, then $r_m \sec \phi = 1$. The angle at this point is precisely equal to $\cos^{-1}(r_m)$. Beyond this angle, as was demonstrated with Figure 1D, the classification error is equal to the minority proportion. The relationship is given by

$$\epsilon(\phi) = \begin{cases} \left[\frac{\phi - r_m^2 \tan\phi}{\pi}\right], & \text{if } 0 \le \phi < \cos^{-1}(r_m), \\ m, & \text{if } \cos^{-1}(r_m) \le \phi \le \frac{\pi}{2}, \end{cases} \tag{2}$$

where we use ϵ to denote classification error in general and $\epsilon(\phi)$ denote the classification error given a specific angle, ϕ.

For two dimensions, our assumption that the alternatives are randomly drawn from a uniform distribution over the circle is equivalent to assuming that ϕ is uniformly distributed. Implicit in

this assumption is the assumption that the dimensions are equally salient and that, given m, the minority can be at any orientation in the space. The distribution of ϕ is, therefore,

$$f(\phi) = \begin{cases} \frac{2}{\pi} & 0 \leq \phi \leq \frac{\pi}{2}, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Given the assumptions of our model and equations (1), (2), and (3), the likelihood that a sample of votes was drawn from a two-dimensional space can be calculated. Specifically, suppose we have a sample of size q. Let m be the observed minority proportion and ϵ_j be the classification error of the j^{th} vote. Given m, equation (1) can be used to solve for r_{mj}; ϵ_j and r_{mj} can be used in equation (2) to solve for ϕ_j; and ϕ_j and $\cos^{-1}(r_{mj})$ can be used in equation (3) to solve for the likelihood. Let $h(\phi_j)$ denote the likelihood for the j^{th} vote. Given m_j and ϕ_j, $h(\phi_j)$ is

$$h(\phi_j) = \begin{cases} \frac{2}{\pi}, & 0 \leq \phi_j < \cos^{-1}(r_{mj}), \\ \frac{2}{\pi} \int_{\cos^{-1}(r_{mj})}^{\frac{\pi}{2}} d\phi = \frac{2}{\pi} \left[\frac{\pi}{2} - \cos^{-1}(r_{mj})\right], & \cos^{-1}(r_{mj}) \leq \phi_j \leq \frac{\pi}{2}. \end{cases} \tag{4}$$

The likelihood function for the two-dimensional case is, therefore,

$$L = \left(\frac{2}{\pi}\right)^q \prod_{j=1}^{q} \Psi_j, \tag{5}$$

where

$$\Psi_j = \begin{cases} 1, & 0 \leq \psi_j < \cos^{-1}(r_{mj}), \\ \left[\frac{\pi}{2} - \cos^{-1}(r_m)\right], & \cos^{-1}(r_{mj}) \leq \psi_j \leq \frac{\pi}{2}. \end{cases}$$

Although the counterparts to equations (1), (2), and (3), are nontrivial for more than two dimensions (equations (A.7), (A.6), and (A.8) in the appendix, respectively), the likelihood function over ϕ for more than two dimensions is conceptually straightforward. Because the sample observations we are interested in are m_j and ϵ_j *not* m_j and ϕ_j pairs, what we need is the likelihood function over ϵ, not ϕ. However, equation (4) is very useful for calculating the point masses which carry over to the likelihood function over ϵ.

For $N = 2$, regardless of the value for m, the density over ϕ is uniform. This is not the case for the density over ϵ. To see this, let $\epsilon = Y(\phi)$ stand for equation (2). Therefore, solving for ϕ, we get $\phi = y^{-1}(\epsilon)$. Using the standard univariate change of variables formula, we can write the density over ϵ as

$$g(\epsilon) = \begin{cases} f\left[y^{-1}(\epsilon)\right] \left|\frac{d(y^{-1}(\epsilon))}{d\epsilon}\right|, & 0 \leq \epsilon < m, \\ \frac{2}{\pi}\left[\frac{\pi}{2} - \cos^{-1}(r_m)\right], & \epsilon = m, \end{cases} \tag{6}$$

where the point mass from equation (4) maps onto m. Unfortunately, inspection of equation (2) shows that $y^{-1}(\epsilon)$ cannot be solved analytically. Consequently, we must derive $g(\epsilon)$ numerically. Figure 3 graphs equations (4) and (6) for an m of 0.35.

The mapping of ϕ to ϵ is one to one but non-linear. Note that as ϕ approaches the critical value, $\cos^{-1}(r_m)$, $g(\epsilon)$ rises rapidly near m. Because fixing m determines r_m, decreasing m increases r_m and decreases the critical value, $\cos^{-1}(r_m)$. Since $h(\phi)$ is uniform with value of $\frac{2}{\pi}$ for values of ϕ less than $\cos^{-1}(r_m)$ regardless of the value of m, this has the effect of increasing the point mass as m decreases. In terms of $g(\epsilon)$, as m decreases—even though the point mass increases—the likelihood values increase over $0 \leq \epsilon < m$. For example, when $m = 0.50$, $g(\epsilon)$ is a uniform distribution with value 2.0. As shown in Figure 3, $g(\epsilon)$ is always greater than 2.120 for $m = 0.35$. For $m = 0.25$, $g(\epsilon)$ is always greater than 2.390; and so on.

B. The General Case

Our assumptions for the general case mirror those for two dimensions. We assume that voters are uniformly distributed through a unit hypersphere, that all votes are binary, and that the N dimensions are equally salient. The minority, m, is the volume defined by the distance, r_m,

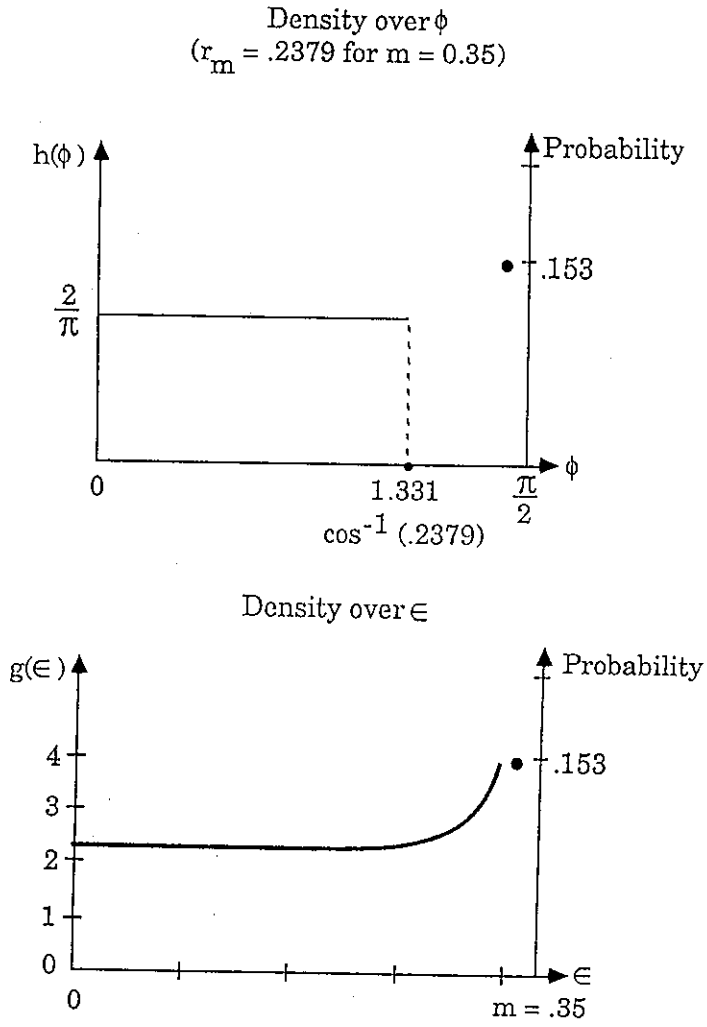

Figure 3. Two-dimensional projection with margin equal to 0.35.

from the origin of the hypersphere to the hyperplane which separates the two choices. We will refer to this hyperplane as the *boundary plane*. Without loss of generality, the projection dimension is assumed to pass through the center of the hypersphere. The angle between the projection dimension and the vector normal to the boundary plane which passes through the center of the hypersphere is denoted as ϕ. The cutting point is that point at which the projection dimension intersects the boundary plane. As with the two-dimensional case, this point is given by $\cos^{-1}(r_m)$. Finally, the error is the sum of two "orange slice" volumes in N-space.

Equation (A.7) in the appendix—the counterpart to equation (1) above—gives the relationship between m and r_m. The complexity of the equation obscures an important geometric fact about this relationship; namely, if m is fixed, then as N increases r_m decreases (see Appendix). This is due to the fact that, as N increases, proportionately more of the volume is concentrated near the center of the unit hypersphere. Hence, if r_m were a fixed constant, then m would fall as N increased.

The distribution of ϕ for $N > 2$ is given in equation (A.8). It is the sine of ϕ taken to the $N-2$ power and multiplied by a constant. As shown in equation (4), the point mass which is common to both $h(\phi)$ and $g(\epsilon)$ is found by taking the integral of $f(\phi)$ from $\cos^{-1}(r_m)$ to $\frac{\pi}{2}$. Intuitively, it is clear that, as N increases, the probability that $\epsilon = m$ increases. Hence, as N increases, the value of the point mass must increase. We prove in the appendix that as N gets large, the point mass converges to $1 - 2m$.

As was the case for $N = 2$, we must derive $g(\epsilon)$ numerically for $N > 2$.[5] Returning to Figure 3, fix m and consider the situation for $N > 2$. If ϵ is very close to zero, then the likelihood value for $N = 2$ must be greater than $N = 3$ and so on in order of N; that is, $g_n(\epsilon) > g_{n+1}(\epsilon)$. Conversely, for an observed error very close to m the likelihood value for $N = 2$ will be smaller than $N = 3$ but, as long as the error is not exactly equal to m, there will be a value of N such that $g_n(\epsilon) > g_{n-1}(\epsilon)$ and $g_n(\epsilon) > g_{n+1}(\epsilon)$. This is so because the point mass increases in N.

Another property that follows from the point masses increasing in N, is the fact that, as N increases, $g(\epsilon)$ must become very small over most of its range and then rise very rapidly near m. Indeed, for large N, $g(\epsilon)$ approaches a right-angle "J" shape. These properties are illustrated by Figure 4.

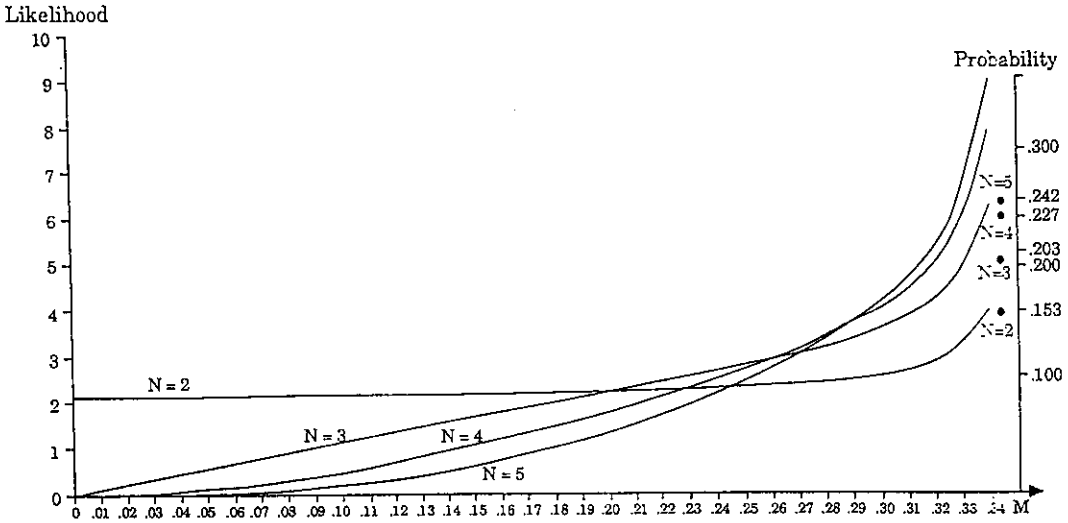

Figure 4. $g(\epsilon)$ functions for $m = .35$.

Figure 4 shows $g(\epsilon)$ for $m = 0.35$ and N equal to 2, 3, 4 and 5. The fact that $g_2(\epsilon) > g_3(\epsilon) > \cdots > g_n(\epsilon)$ for ϵ sufficiently close to zero, and the fact that $g_n(\epsilon) > g_{n-1}(\epsilon) > \cdots > g_2(\epsilon)$ for ϵ sufficiently close to m, means that between $0 \leq \epsilon < m$, $g_n(\epsilon)$ must cross all the other densities, $g_{n-1}(\epsilon)$ must cross those densities for $N - 2$ and below, and so on.

The crossing points of the $g(\epsilon)$'s can be clearly seen in Figure 4. Let ϵ_{nn+1} denote the value of ϵ such that $g_n(\epsilon) = g_{n-1}(\epsilon)$ and let g_{nn+1} denote the value $g_n(\epsilon)$ corresponding to ϵ_{nn+1}. The properties of $g(\epsilon)$ discussed above imply that $\epsilon_{23} < \epsilon_{24} < \cdots < \epsilon_{2n}$, $\epsilon_{34} < \epsilon_{35} < \cdots < \epsilon_{3n}$, and so on; and that $g_{23} < g_{24} < \cdots < g_{2n}$, $g_{34} < g_{35} < \cdots < g_{3n}$, and so on.

Figure 4 also shows the important fact that, for a fixed m and ϵ, the likelihood values over N are single peaked. This flows directly from the fact that $g(\epsilon)$ is monotonically increasing. Between zero and approximately 0.20, $g_2(\epsilon) > g_3(\epsilon) > g_4(\epsilon) > g_5(\epsilon)$, between approximately 0.20 and approximately 0.23, $g_3(\epsilon) > g_2(\epsilon) > g_4(\epsilon) > g_5(\epsilon)$, and so on.

3. MONTE CARLO TESTS OF THE LIKELIHOOD FUNCTION

In order to obtain our theoretical results, we assumed that the number of voters was infinite so that we could treat the binary choices as hypervolumes. As a practical matter, a sample of m and ϵ pairs will be drawn from a policy space with a finite number of voters. This will not greatly affect our results because, if the voters are treated as a random sample from a uniform distribution through the hypersphere, then the larger the number of voters the closer the approximation between the hypervolumes and the corresponding ratios of types of voters

[5] In particular, we computed $g(\epsilon)$ for N's of 2 through 8 for all values of m and ϵ from .025 to .500 in thousandths (i.e., .025, .026, .027,..., .498, .499, .500). That is, we produced four 476 by 476 tables. These tables were computed on the CRAY YMP at the Pittsburgh Supercomputer Center and were used in the Monte Carlo and empirical analyses below. The tables are available form the authors on request.

to the total sample size. However, for a fixed number of voters, the accuracy of the likelihood function must fall with N. Consequently, we performed a Monte Carlo study to determine how sensitive our likelihood function was to various values of the number of voters and the size of the sample of m and ϵ pairs. A portion of our study is shown in Table 1.

Table 1. Monte Carlo tests.

	Number of Votes																	
	100						500						1000					
	True Dimensionality																	
	2*	3	4	5	6	7	2	3	4	5	6	7	2	3	4	5	6	7
	Experiments With 100 Voters																	
2	100#	42	3	0	0	0	100	39	0	0	0	0	100	37	0	0	0	0
3	0	57	83	46	20	6	0	61	95	57	21	4	0	63	93	59	21	4
4	0	1	14	54	75	79	0	0	5	43	74	81	0	0	7	41	74	81
5	0	0	0	0	5	15	0	0	0	0	5	15	0	0	0	0	5	15
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Experiments With 500 Voters																	
2	100	2	0	0	0	0	100	0	0	0	0	0	100	0	0	0	0	0
3	0	96	18	0	0	0	0	100	13	0	0	0	0	100	8	0	0	0
4	0	2	79	62	10	0	0	0	87	64	5	0	0	0	92	65	2	0
5	0	0	3	38	78	42	0	0	0	36	94	45	0	0	0	35	94	52
6	0	0	0	0	12	51	0	0	0	0	1	55	0	0	0	0	4	48
7	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Experiments With 1000 Voters																	
2	100	1	0	0	0	0	100	0	0	0	0	0	100	0	0	0	0	0
3	0	96	8	0	0	0	0	100	1	0	0	0	0	100	0	0	0	0
4	0	0	86	22	1	0	0	0	99	22	0	0	0	0	100	8	0	0
5	0	0	6	72	53	10	0	0	0	78	54	2	0	0	0	92	35	1
6	0	0	0	6	42	72	0	0	0	0	46	91	0	0	0	0	65	92
7	0	0	0	0	4	17	0	0	0	0	0	7	0	0	0	0	0	7
8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

* For each combination of Voters and Votes, the columns give the distribution of 100 samples taken from a space with the indicated dimensionality across values of N.

\# Each cell entry is the number of samples with maximum likelihood at the row value of N.

Table 1 consists of 9 cells with 600 experiments per cell for a total of 5400 experiments. Each of the 9 cells is a combination of a number of voters—100, 500, and 1000—and a number of votes—again, 100, 500, and 1000. For each combination of voters and votes we drew 100 random samples from 2-, 3-, 4-, 5-, 6-, and 7-dimensional spaces for a total of 600 random samples. For each sample we computed the likelihood that it was drawn from a 2-, 3-, 4-, 5-, 6-, 7-, and 8-dimensional space. Each entry in a column (the true N) within one of the 9 cells of Table 1 is the number of samples with a maximum likelihood corresponding to the row value for N. For example, consider the 100 random samples with 500 voters and 500 votes taken from a space with $N = 4$. None of the 100 random samples had a maximum likelihood which peaked at $N = 2$, thirteen of the 100 samples peaked at $N = 3$, and eighty-seven peaked at the true value $N = 4$.[6]

[6]The voters were drawn from a uniform distribution through the hypersphere and the voters were created by

The results in Table 1 show that the number of voters in the sample is more important than the number of votes. Holding fixed the number of voters and doubling the number of votes from 500 to 1000 has very little effect. In contrast, holding the number of votes fixed and doubling the number of voters from 500 to 1000 has a much greater effect. As expected, with smaller numbers of voters there is a distinct downward bias—the likelihood function tends to peak at a value one less than the true value.

4. AN APPLICATION TO ROLL CALL VOTING IN THE U.S. CONGRESS

Given the quality of our Monte-Carlo results with 500 voters and above, a natural application for our likelihood function is the recorded roll call votes in the U.S. House of Representatives because it has had over 400 members since 1903. We obtained two ϵ vectors for testing. The first is from Poole and Rosenthal's [1991] dynamic one-dimensional scaling of the House roll calls from 1789 to 1985 (Congresses 1 through 99). Poole and Rosenthal's [1989, 1991] D-NOMINATE technique maximizes a likelihood function—it does not minimize classification error. In addition, the legislator coordinates are constrained to be simple polynomial functions of time. Because of this, the ϵ vector from the D-NOMINATE one-dimensional scaling will tend to be inflated. In particular, it is possible for an observed ϵ to be greater than m for a roll call. If this occurred, we set $\epsilon = m$.

We obtained the second ϵ vector by performing a simple Guttman scaling of each matrix of House roll call votes. The ϵ vector from the Guttman scaling will tend to be smaller than that from D-NOMINATE because Guttman scaling minimizes classification errors and each Congress was scaled separately.

Before turning to the results, a word of caution is necessary. Our approach assumes that the the m and ϵ vectors are random draws. That is, each pair of m and ϵ elements in the sample is independent and identically distributed. The observed m and ϵ vectors from various scaling technologies do not possess this property. This is because any scaling/projection technique must work with the rectangular matrix (voters by votes) of binary choices. Typically the similarity between the voting records of the individuals (similarity between rows) is used to place the individuals on a dimension and the similarity between the columns is used to place the votes. Clearly, changing a column will affect the placement of all the individuals and changing a row will affect the placement of all the votes. If the individual placements change, then the cutting points could change thereby changing some of the ϵ's. Hence, the ϵ's are not independent. Furthermore, any scaling/projection technique will—much like eigenvalue-eigenvector decomposition—tend to place the first dimension through the major axis of the point cloud. Consequently, if the ϵ's are obtained by projecting onto this first dimension, then the ϵ's will be biased downward.

This dependency problem is inherent in any empirical method of projection and the seriousness of the problem is a function of sample size. The larger the number of voters/votes, the less the effect from changing one voter/vote and the closer the observed ϵ will be to the true ϵ. The relationship between the sample size and the degree of dependence of the errors is outside the scope of this paper.

Figure 5 graphs the geometric mean likelihoods (i.e., the log likelihood divided by the number of votes and exponentiated) for $N = 2$ and $N = 3$ for all 99 Congresses calculated from the two ϵ vectors. To reduce the clutter in the figure, we show the graphs for $N = 4$ and $N = 5$ only when they are greater than either $N = 2$ or $N = 3$.

The striking fact about Figure 5 is the clear dominance of $N = 2$. For every Congress since 1853, the likelihood function peaks at $N = 2$. The likelihood of $N = 3$ becomes closer to $N = 2$ after World War I and almost surpasses it during the 85th Congress. Prior to 1853 $N = 2$ predominates except for the Era of Good Feeling (approximately the 14^{th} through the

randomly drawing the policy outcome pairs from within the hypersphere. The first dimension of the space was used as the projection dimension and for each artificial vote, ϵ was found by testing every possible cutting point and utilizing the point which minimized ϵ. Finding the best possible point is easy. The number of possible cutting points is equal to the number of voters. To see this, let p be the number of voters. The cutting point could be between any adjacent pair of voters for $p - 1$ points, and at the end of the projection dimension for the p^{th} point. For the final case, $\epsilon = m$.

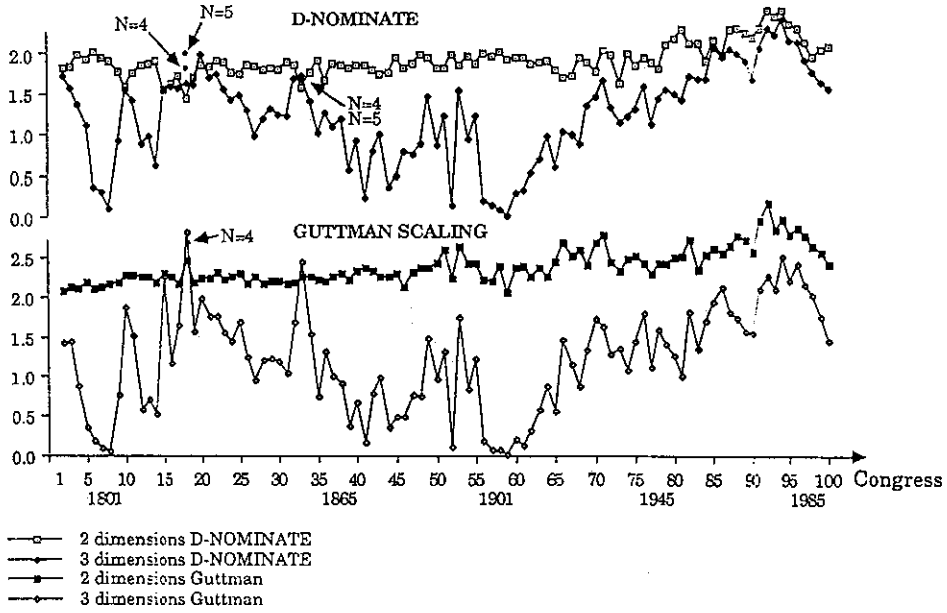

Figure 5. Likelihoods for U.S. HOUSE.

19th Congresses) and the Congress after the compromise of 1850.[7] However, the size of the House in the 18th and early 19th centuries varied from 66 members to around 200 members so these earlier results cannot be judged to be as reliable as those for the 20th century.

The graphs for the likelihoods given by the ϵ's from the two different scaling techniques are almost identical. Given that it is reasonable to assume that the ϵ's from D-NOMINATE are inflated and those from the Guttman scaling are deflated, then the likelihoods computed using the ϵ's from D-NOMINATE should be biased upwards in terms of N while those from the Guttman scaling should be biased downwards. Indeed, if we included the graphs for $N = 4$ and $N = 5$ in Figure 5 then this effect shows up in three of the earlier Congresses—the 17th, 19th, and 32nd. For example, the order of the likelihoods for the 17th Congress was 5-4-3-2 using the D-NOMINATE ϵ's; while using the Guttman scaling ϵ's the order was 3-4-2-5.

5. CONCLUSION

The purpose of this paper was to develop a new approach to the problem of determining the dimensionality of a spatial voting model. Determining the dimensionality of a matrix of binary choice data with traditional scaling techniques (e.g., factor analysis or multi-dimensional scaling) is difficult because these techniques use measures of fit that always improve with the number of dimensions being estimated. A decision that k is the correct number of dimensions is really a decision about the significance of the *rate of the rate of change* of the measure of fit—namely, the increase in fit from $k-1$ to k dimensions is more "significant" than the increase in fit from k to $k+1$ dimensions. This decision is further complicated by the fact that the number of parameters must increase with N because an additional coordinate for each individual must be estimated for each additional dimension.

Our approach is independent of the estimation method. We assume a uniform distribution of voters through an N-dimensional unit hypersphere. We assume perfect spatial voting with equally salient dimensions and solve for the projection of this perfect voting onto one dimension and the *resultant classification* error. We derive the probability density function over the classification errors which can be used to calculate the likelihood that a sample of votes was drawn from an N-dimensional hypersphere. Our Monte-Carlo investigation into the properties of this likelihood function shows that it can be used reliably for low N with small sample sizes of both voters and votes. For $N \geq 4$, somewhat larger sample sizes of voters insure greater reliablity.

[7] For a substantive discussion of the apparent multi-dimensionality of roll call voting during these periods, see Poole and Rosenthal [1989, 1991].

The likelihood function we have developed can serve as a useful null model for evaluating estimated spatial models. The results of our empirical application to roll call voting in the U.S. House of Representatives, support the recent body of empirical work on roll call voting which argues for low dimensionality.

References

Austen-Smith, D., The spatial theory of electoral competition: Instability, Institutions, and Information, *Environment and Planning C: Government and Policy* **1**, 439–459 (1983).

Austen-Smith, D. and J. Banks, Electoral accountability and incumbency, In *Models of Strategic Choice in Politics*, (Edited by P. Ordeshook), University of Michigan Press, Ann Arbor, (1989).

Austen-Smith, D. and J. Banks, Stable governments and the allocation of policy portfolios, *American Political Science Review* **84**, 891–906 (1990).

Banks, J., A model of electoral competition with incomplete information, *Journal of Economic Theory* **50**, 309–325 (1990).

Black, D., *The Theory of Committees and Elections*, Cambridge University Press, Cambridge, (1958).

Cahoon, L., Locating a set of points using range information only, Ph.D. Dissertation, Carnegie-Mellon University, Pittsburgh, PA, (1975).

Calvert, R., Robustness of the multi-dimensional voting model: Candidate motivations, uncertainty and convergence, *American Journal of Political Science* **29**, 69–95 (1985).

Converse, P.E., The nature of belief systems in mass publics, In *Ideology and Discontent*, (Edited by D. E. Apter), Free Press, New York, (1964).

Downs, A., *An Economic Theory of Democracy*, Harper & Row, New York, (1957).

Enelow, J. and M. Hinich, *The Spatial Theory of Voting*, Cambridge University Press, New York, (1984).

Hinich, M.J. and W. Pollard, A new approach to the spatial theory of electoral competition, *American Political Science Review* **25**, 323–341 (1981).

Koford, K., Dimensions in congressional voting, *American Political Science Review* **83**, 949–962 (1989).

Krehbiel, K., Spatial models of legislative choice, *Legislative Studies Quarterly* **13**, 259–319 (1988).

Krehbiel, K. and D. Rivers, The analysis of committee power: An application to Senate voting on the minimum wage, *American Journal of Political Science* **32**, 1151–1174 (1988).

Krehbiel, K. and D. Rivers, Sophisticated voting in congress: A reconsideration, *The Journal of Politics* **52**, 548–578 (1990).

MacRae, D. Jr., *Dimensions of Congressional Voting*, University of California Press, Berkeley, (1958).

MacRae, D. Jr., *Issues and Parties in Legislative Voting*, Harper & Row, New York, (1970).

Ordeshook, P., *Game Theory and Political Theory*, Cambridge University Press, Cambridge, (1986).

Poole, K., Recent developments in analytical models of voting in the U.S. Congress, *Legislative Studies Quarterly* **13**, 117–133 (1988).

Poole, K. and H. Rosenthal, U.S. presidential elections 1968–80: A spatial analysis, *American Journal of Political Science* **28**, 282–312 (1984).

Poole, K. and H. Rosenthal, Color animation of dynamic congressional voting models, GSIA Working Paper no. 64-88-89, Carnegie-Mellon University, Pittsburgh, PA, (1989).

Poole, K. and H. Rosenthal, A dimension is a dimension is a dimension, GSIA Working Paper, February, 1990, (1990).

Poole, K. and H. Rosenthal, Patterns of congressional voting, *American Journal of Political Science* **35**, 228–278 (1991).

Rabinowitz, G., A procedure for ordering object pairs consistent with the multi-dimensional unfolding model, *Psychometrika* **45**, 349–373 (1976).

Rusk, J. and H. Weisberg, Perceptions of presidential candidates: Implications for electoral change, *Midwest Journal of Political Science* **16**, 388–410 (1972).

Weisberg, H. and J. Rusk, Dimensions of candidate evaluation, *American Political Science Review* **64**, 1167–1185 (1970).

APPENDIX

1. Introduction

In this appendix we derive the relationships between ϵ and ϕ, m and ϕ, and the likelihood functions, $h(\phi)$ and $g(\epsilon)$ for all N. We also show proofs that the cutting plane (defined below) minimizes the classification error; that, with fixed m, as N increases, r decreases; and that as N increases, the point masses converge to $1 - 2m$.

We assume that the distribution of voters is uniform through the N-dimensional unit sphere. Choices are binary so that the N-dimensional sphere is divided into two regions on every vote. The minority is represented by the upper m percent of the volume of the sphere and the majority is represented by the lower $1 - m$ percent of the

hypersphere. We refer to the hyperplane that separates the two regions as the *boundary plane* and we refer to the axis that is orthogonal to the boundary plane as the *z-axis*. The distance up the z-axis from the origin to the boundary plane is denoted as r_m.

Without loss of generality, the classification dimension is fixed and passes through the center of the hypersphere. For a fixed m, the classification error is a function only of the angle, ϕ, between the z-axis of the hypersphere and the classification dimension.

2. Derivation of $\epsilon(\phi)$

The basic tool of our derivation is the uniform density through the N-dimensional unit sphere, parameterized in polar coordinates. This can be achieved by starting with a uniform density in a N-dimensional rectangular coordinate system and then transforming it to N-dimensional polar coordinates.

A point $(r, \alpha_1, \alpha_2, \ldots, \alpha_{N-2}, \theta)$ in the N-dimensional polar coordinate system, where

$$0 \leq r, \quad 0 \leq \alpha_1 \leq \pi, \ldots, 0 \leq \alpha_{N-2} \leq \pi, \quad 0 \leq \theta \leq 2\pi,$$

is mapped to the point $(y_1, y_2, \ldots, y_N)$ in the N-dimensional rectangular coordinate system by the transformation:

$$\begin{aligned}
y_1 &= r\cos\theta \sin\alpha_1 \sin\alpha_2 \ldots \sin\alpha_{N-3} \sin\alpha_{N-2} \\
y_2 &= r\sin\theta \sin\alpha_1 \sin\alpha_2 \ldots \sin\alpha_{N-3} \sin\alpha_{N-2} \\
y_3 &= r \quad\ \cos\alpha_1 \sin\alpha_2 \ldots \sin\alpha_{N-3} \sin\alpha_{N-2} \\
&\vdots \qquad\qquad\quad \vdots \\
y_{N-1} &= r \qquad\qquad\qquad \cos\alpha_{N-3} \sin\alpha_{N-2} \\
y_N &= r \qquad\qquad\qquad\qquad \cos\alpha_{N-2}.
\end{aligned} \tag{A.1}$$

The Jacobian, J_N, of this transformation is

$$J_N = r^{N-1}\sin(\alpha_1)\sin^2(\alpha_2)\ldots\sin^{N-2}(\alpha_{N-2}). \tag{A.2}$$

The volume and surface area of the N-dimensional unit sphere will be denoted by V_N and S_N, respectively. The equations for V_N and S_N are

$$V_N = \frac{\pi^{N/2}}{\Gamma\left(\frac{N}{2}+1\right)} \tag{A.3}$$

$$S_N = \frac{N\pi^{N/2}}{\Gamma\left(\frac{N}{2}+1\right)}. \tag{A.4}$$

When using polar coordinates, it is critical that the model be correctly oriented. With N-dimensional polar coordinates there are three "types" of dimensions which must be considered. The first ranges from 0 to $\frac{2}{\pi}$ and will be denoted by θ. This is the angle which sweeps around the classification dimension. The second, which is the radius, r, ranges from 0 to 1. Finally, the remaining $N-2$ dimensions each range from 0 to π and will be denoted $\alpha_1, \alpha_2, \ldots, \alpha_{N-2}$. The ordering of these $N-2$ dimensions is at times important. If this is the case, then the "last" of these $N-2$ dimensions will be denoted by ϕ.

Because the classification error, $\epsilon(\phi)$, is a function only of the angle between the z-axis and the classification dimension, the general problem reduces to looking at cross sections containing the classification dimension and the z-axis. By symmetry attention can be limited to $\phi \in (0, \frac{\pi}{2})$.

The point where the classification dimension intersects the boundary plane of the hypersphere will be designated the *classification point*. The *cutting plane* is the hyperplane that passes through the classification point and is orthogonal to the classification dimension. The classification error is the percent of the hypersphere's volume that is bounded between the boundary plane and the classification plane. The cutting plane minimizes the classification error. All other hyperplanes orthogonal to the classification dimension but not passing through the classification point will produce a larger classification error than the cutting plane. We prove this below.

If $\cos^{-1}(r_m) \leq \phi \leq \frac{\pi}{2}$, the classification error is the entire minority volume. This corresponds to the error being minimized by classifying all voters into the majority; that is

$$\epsilon(\phi) = m \qquad \cos^{-1}(r_m) \leq \phi \leq \frac{\pi}{2}.$$

If $0 \leq \phi < \cos^{-1}(r_m)$, then the classification error is the sum of two volumes. One volume represents voters in the minority that are incorrectly classified in the majority and the other volume represents voters in the majority that are incorrectly classified in the minority. One way to calculate the percent of the volume of the hypershpere that is misclassified is to switch the volume associated with the misclassified majority with a comparable volume in the minority. This highlights the fact that the classification error must always be less than or equal to m. To

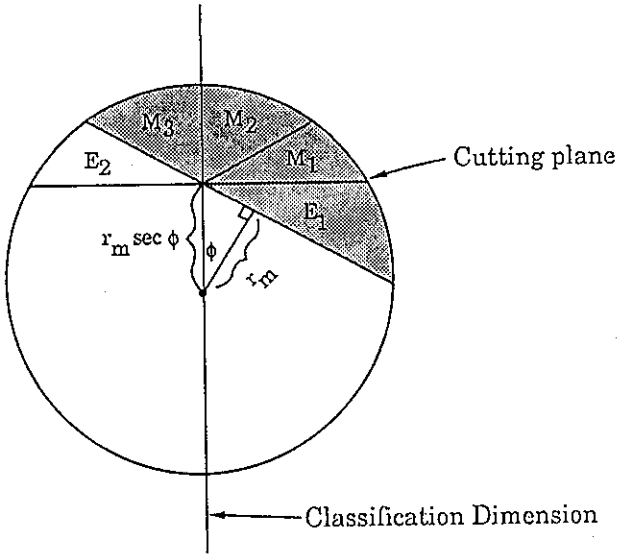

Figure A.1. Error and minority regions.

calculate the classification error, subtract from m the percent of the volume of the hypersphere that corresponds to the remaining correctly classified volume in the minority region.

For example, Figure A.1—which duplicates Figure 2 in the text—shows what this process looks like for two dimensions. The total classification error is the sum of E_1 and E_2 and the minority region is the sum of E_1, M_1, M_2, and M_3. Region M_1, which is in the correctly classified minority region, is the same size as E_2. Consequently, if we switch regions E_2 and M_1 then the classification error is equal to $m - 2M_2$ because $M_2 = M_3$. Hence our problem reduces to finding the general formula for the volume of M_2 for any N.

This volume can be indexed along the dimension in the cross section that is orthogonal to the classification dimension and passes through the origin. Figure A.2 shows how this is accomplished. In Figure A.2 we label the axes of the cross section x and y respectively. This enables us to write the equation of the line bordering M_2 as $y = r_m \sec\phi + (\tan\phi)x$. With this line equation it is simple to solve for the coordinates of the point of intersection of the line with the circle. In particular, the x axis coordinate is $\cos\phi\sqrt{1 - r_m^2} - r_m \sin\phi$.

We now can write the limits for the integral which will yield the required volume corresponding to M_2. In particular, let ρ_N be an index which ranges from 0 to $\cos\phi\sqrt{1 - r_m^2} - r_m \sin\phi$. At each point, ρ_N, of this index (the "x" dimension in Figure A.2) the cross section through the N-dimensional sphere is an $N - 1$-dimensional sphere with radius $\sqrt{1 - \rho_N^2}$ (the "y" dimension in Figure A.2). The volume associated with the correctly classified region is the dome above $r_m \sec\phi + (\tan\phi)\rho_N$. Hence, to obtain the required volume, we integrate ρ_N from 0 to $\cos\phi\sqrt{1 - r_m^2} - r_m \sin\phi$; and, for each ρ_N, we integrate from $r_m \sec\phi + (\tan\phi)\rho_N$ to $\sqrt{1 - \rho_N^2}$. The entire result is then multiplied by 2 to obtain the volume associated with $2M_2$.

Our problem therefore reduces to the calculation of the upper dome of an $(n-1)$-dimensional sphere. This is a simple problem in geometry. Let the radius of the $(N-1)$-dimensional sphere be r. The volume of the dome that is s units above the equator is given by

$$V_{N-2} \int_s^r \left(\sqrt{r^2 - \rho_{N-1}^2}\right)^{N-2} d\rho_{N-1} \tag{A.5}$$

The volume in the N-dimensional sphere that corresponds to the classification error is m minus the percent of the hypersphere's volume that is in the minority region. Using equation (A.5) and the index we developed above, the classification error function, $\epsilon(\phi)$, for $0 \leq \phi < \cos^{-1}(r_m)$ can be written as

$$\epsilon(\phi) = m - \frac{2}{V_N} \int_0^{\cos\phi\sqrt{1-r_m^2} - r_m \sin\phi} V_{N-2} \int_{r_m \sec\phi + \rho_N \tan\phi}^{\sqrt{1-\rho_N^2}} \left(\sqrt{1 - \rho_N^2 - \rho_{N-1}^2}\right)^{N-2} d\rho_{N-1}\, d\rho_N$$

simplifying

$$\epsilon(\phi) = \begin{cases} m - \frac{N}{\pi} \displaystyle\int_0^{\cos\phi\sqrt{1-r_m^2} - r_m \sin\phi} \int_{r_m \sec\phi + \rho_N \tan\phi}^{\sqrt{1-\rho_N^2}} \left(\sqrt{1 - \rho_N^2 - \rho_{N-1}^2}\right)^{N-2} d\rho_{N-1}\, d\rho_N, \\ \qquad\qquad \text{for } 0 \leq \phi < \cos(r_m^{-1}) \\ m, \qquad\qquad \text{for } \cos^{-1}(r_m) \leq \phi \leq \frac{\pi}{2}. \end{cases} \tag{A.6}$$

Equation (A.6) is a generalization of equation (2) in the text.

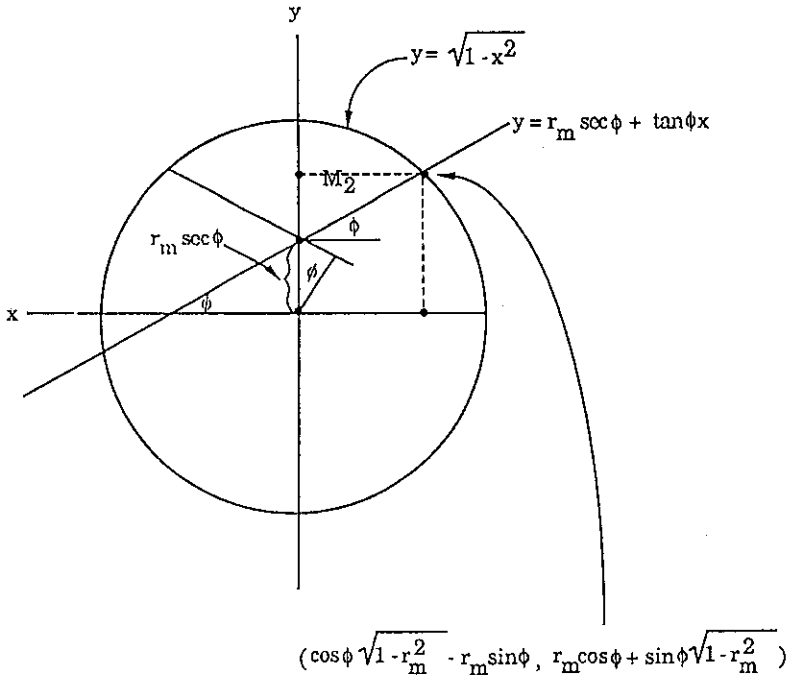

Figure A.2. Obtaining limits of integration.

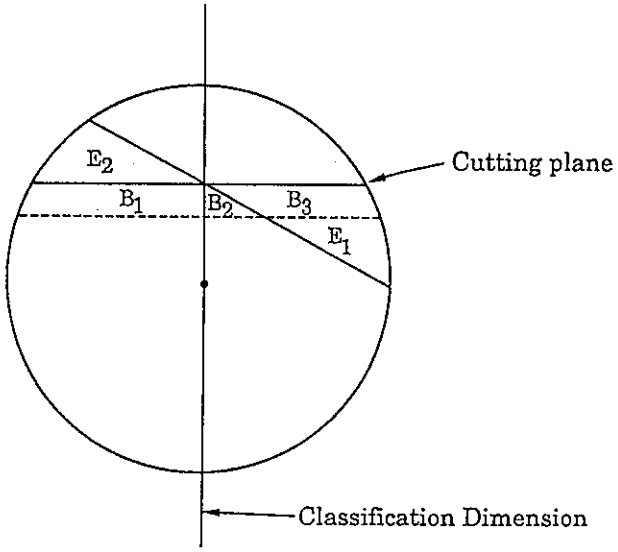

Figure A.3. Cutting plane demonstration.

The index, ρ_N, allows us to write the minority percentage, m, as a simple integral. The minority percentage is the volume of the dome above the boundary plane of the N-dimensional sphere divided by the total volume. Using (A.5) the formula for m is

$$m = \frac{V_{N-1}}{V_N} \int_{r_m}^{1} \left(\sqrt{1 - \rho_N^2}\right)^{N-1} d\rho_N \tag{A.7}$$

Equation (A.7) is the generalization of equation (1) in the text.

3. A Proof that the Cutting Plane Minimizes Classification Error

In the above development we assumed that the cutting plane—the hyperplane that passes through the classification point and is orthogonal to the classification dimension—minimizes the classification error. We now offer a simple geometric proof that the cutting plane does minimize the classification error.

THEOREM A.1. *The Cutting Plane Minimizes the Classification Error*

PROOF. Consider Figure A.3 which is a simplified version of Figure A.1. The sum of the volumes associated with regions E_1 and E_2 is the classification error. Now consider an alternative hyperplane orthogonal to the classification dimension which we represent with a dotted line in Figure A.3. By symmetry, the volume associated with region $B1$ must equal the volume assocated with region $B2+B3$. Hence, if the alternative hyperplane is used to calculate classification error then the volume associated with region $B3$ will switch from being incorrectly classified to being correctly classified while the volumes associated with regions $B1$ and $B2$ will switch from being correctly classified to being incorrectly classified. However, since the volume associated with $B1 + B2$ is greater than the volume associated with $B3$, the alternative hyperplane is worse than the cutting plane. A similar demonstration can be given for hyperplanes above the cutting plane in Figure A.3. ∎

4. Derivation of $f(\phi)$, $h(\phi)$, and $g(\epsilon)$

We assume that the two alternative points on any vote are randomly drawn from within the N-dimensional sphere. Implicit in this assumption is the assumption that the N dimensions are equally salient and that, given m, the minority volume (upper dome) can be at any orientation vis a vis the classification dimension. In two dimensions this is equivalent to the assumption that ϕ has a uniform distribution. In N dimensions this is equivalent to the assumption that the z-axis can pass through any point on the surface of the sphere. To obtain the uniform density over the surface of the N-dimensional unit sphere, we use polar coordinates as given in equations (A.1) and (A.2) and we set $\alpha_{N-2} = \phi$. By symmetry, we limit attention to the range $0 < \theta < \frac{\pi}{2}$ and multiply the result by 8. Using the equation for the surface area, (A.4), this yields

$$\begin{aligned} f(\phi) &= \frac{8}{S_N} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{\frac{\pi}{2}} |J_N|_{r=1} \, d\theta d\alpha_1 \, d\alpha_2 \, d\alpha_3 \ldots d\alpha_{N-3} \\ &= \frac{8}{S_N} (\sin\phi)^{N-2} \int_0^{\pi} \cdots \int_0^{\pi} |J_{N-1}|_{r=1} \, d\alpha_1 \, d\alpha_2 \, d\alpha_3 \ldots d\alpha_{N-3} \int_0^{\frac{\pi}{2}} d\theta \\ &= \frac{8}{S_N} (\sin\phi)^{N-2} \frac{S_{N-1}}{2\pi} \frac{\pi}{2} \end{aligned}$$

simplifying

$$f(\phi) = \frac{2S_{N-1}}{S_N} (\sin\phi)^{N-2}, \qquad 0 \le \phi \le \frac{\pi}{2}. \tag{A.8}$$

Equation (A.8) is the generalization of equation (3) in the text.

The likelihood function, $h(\phi)$, follows immediately from equation (A.8); namely

$$h(\phi) = \begin{cases} \frac{2S_{N-1}}{S_N} (\sin\phi)^{N-2} & 0 \le \phi < \cos^{-1}(r_m), \\ \frac{2S_{N-1}}{S_N} \int_{\cos^{-1}(r_m)}^{\frac{\pi}{2}} (\sin x)^{N-2} \, dx & \cos^{-1}(r_m) \le \phi \le \frac{\pi}{2}. \end{cases} \tag{A.9}$$

To derive $g(\epsilon)$, let $\epsilon = y(\phi)$ stand for equation (A.6). Solving for ϕ, we get $\phi = y^{-1}(\epsilon)$. Using the univariate change of variables formula, we can write the density over ϵ as

$$g(\epsilon) = \begin{cases} f\left[y^{-1}(\epsilon)\right] \left| \frac{d(y^{-1}(\epsilon))}{d\epsilon} \right|, & 0 \le \epsilon < m, \\ \frac{2S_{N-1}}{S_N} \int_{\cos^{-1}(r_m)}^{\frac{\pi}{2}} (\sin x)^{N-2} \, dx, & \epsilon = m. \end{cases} \tag{A.10}$$

Equation (A.10) is the generalization of equation (6) in the text.

5. A Proof that with Fixed m, as N Increases, r_m Decreases

To avoid confusion, r_m will be superscripted to denote its dependence on N.

THEOREM A.2. *For fixed m, as N increases r_m^N decreases monotonically.*

PROOF. The value r_m^N is defined by

$$m = \frac{V_{N-1}}{V_N} \int_0^{\cos^{-1}(r_m^N)} \sin^N \phi d\phi$$

Define the function: $h_n(\phi) = \frac{V_{N-1}}{V_N} \sin^N \phi$. The functional form of $h_N(\phi)$ implies:

1. for all N $\int_0^{\frac{\pi}{2}} h_N(\phi) \, d\phi = \frac{1}{2}$,
2. for $N \neq N'$, $h_N(\phi) = h_N$, (ϕ) at only one point in $(0, \frac{\pi}{2})$, and
3. for $N > N'$, $h_N(\frac{\pi}{2}) > h_N$, $(\frac{\pi}{2})$.

Hence, for $N > N'$ and any $\theta \in (0, \frac{\pi}{2})$,

$$\int_0^\theta [h_N(\phi) - h_{N'}(\phi)]\, d\phi < 0,$$

which implies $\cos^{-1}(r_m^N) > \cos^{-1}(r_m^{N'})$ or alternatively that $r_m^N < r_m^{N'}$ ∎

6. A Proof that as N Increases, the Point Mass Converges to $1 - 2m$

THEOREM A.3. *For fixed* m, *the limit as* N *increases of the point masses is* $1 - 2m$.

PROOF. For fixed m and N the point mass is given by

$$\begin{aligned}
&\frac{2S_{N-1}}{S_N} \int_{\cos^{-1}(r_m^N)}^{\frac{\pi}{2}} \sin^{N-2}\phi d\phi \\
&= \frac{2S_{N-1}}{S_N} \left[\int_0^{\frac{\pi}{2}} \sin^{N-2}\phi d\phi - \int_0^{\cos^{-1}(r_m^N)} \sin^{N-2}\phi d\phi \right] \\
&= \frac{2S_{N-1}}{S_N} \int_0^{\frac{\pi}{2}} \sin^{N-2}\phi d\phi - \frac{2(N-1)}{N} \frac{V_{N-1}}{V_N} \int_0^{\cos^{-1}(r_m^N)} \sin^{N-2}\phi d\phi \\
&= 1 - \frac{2(N-1)}{N} \frac{N}{(N-1)} \left[m + \frac{1}{N} \frac{V_{N-1}}{V_N} r_m^N \left(\sqrt{1 - (r_m^N)^2} \right)^{N-1} \right] \\
&= 1 - 2m - \frac{2}{N} \frac{V_{N-1}}{V_N} r_m^N \left(\sqrt{1 - (r_m^N)^2} \right)^{N-1}.
\end{aligned}$$

Now, because as $k \to \infty$ $\frac{\Gamma(k+a)}{\Gamma(k+b)} \sim k^{a-b}$, we have as $N \to \infty$ $\frac{V_{N-1}}{V_N} \sim \frac{\sqrt{N}}{\sqrt{2\pi}}$.
Combined with the fact that $r_m^N \leq 1$, this implies that the last term converges to zero. ∎

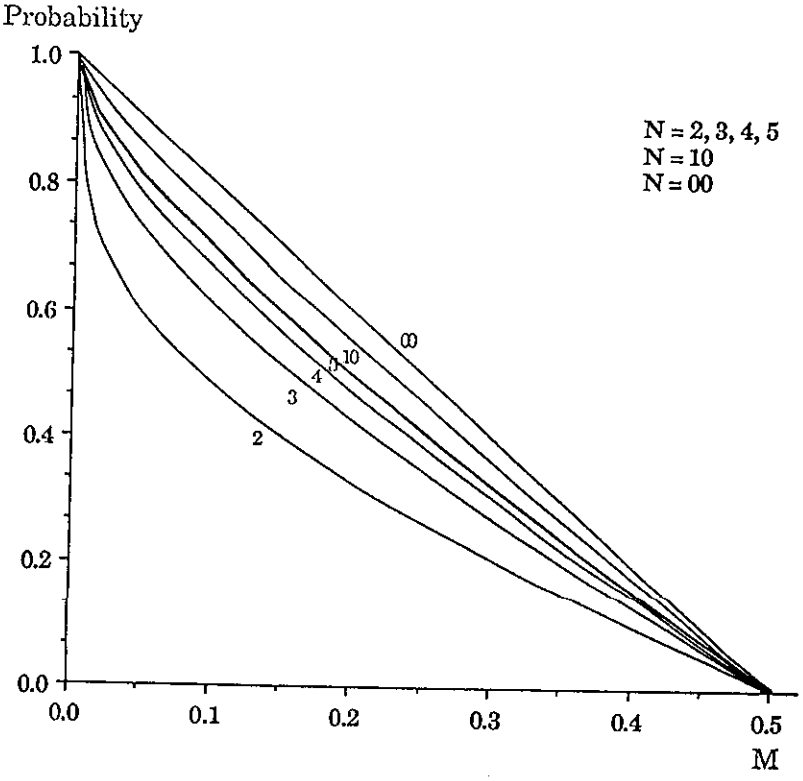

Figure A.4. Point mass values.

7. A Conjecture that the Point Masses Increase Monotonically With N

CONJECTURE A.1. *For fixed m, as N increases the point masses increase Montonically.*

Define the point mass function

$$pt(m,N) = \frac{2S_{N-1}}{S_N} \int_{\cos^{-1}(r_m^N)}^{\frac{\pi}{2}} \sin^{N-2} \phi d\phi$$

This function is plotted for several values of N in Figure A.4.

It appears that for $N > N'$, $pt(m,N) \leq pt(m,N')$ for all m. However, we have not been able to obtain a proof. Extensive searching has failed to produce a counter example.