BAYESIAN METRIC MULTIDIMENSIONAL SCALING

by

Ryan Bakker

Keith T. Poole

Department of Political Science

School of Public and International Affairs

University of Georgia

ABSTRACT

In this paper we show how to apply Bayesian methods to noisy ratio scale distances for both the classical similarities problem as well as the unfolding problem. Bayesian methods produce essentially the same point estimates as the classical methods but are superior in that they provide more accurate measures of uncertainty in the data. Identification is non-trivial for this class of problems because a configuration of points that reproduces the distances is only identified up to a choice of origin, angles of rotation, and sign flips on the dimensions. We prove that fixing the origin and rotation is sufficient to identify a configuration in the sense that the corresponding maxima/minima are inflection points with full rank Hessians. However, an unavoidable result is multiple posterior distributions that are mirror images of one another. This poses a problem for MCMC methods. The approach we take is to find the optimal solution using standard optimizers. The configuration of points from the optimizers is then used to isolate a single Bayesian posterior which can then be easily analyzed with standard MCMC methods.

1. Introduction

In this paper we take a fresh look at the classical ratio scale similarities and unfolding problems from the Psychometrics literature using Bayesian methods. Similarities and unfolding are Multidimensional Scaling (MDS) methods. Multidimensional Scaling encompasses a wide-variety of statistical techniques aimed at characterizing structure within a set of preference or perceptual data. The most common uses of MDS are to uncover the dimensionality of given set of data and to visually display the placements of stimuli (i.e. products, candidates, etc.) according to their positions on the dimension(s) (this is the similarities problem). Additionally, there is often interest in the placement of respondents who have expressed preferences for the stimuli and therefore can be located on the dimension(s) relative to these stimuli (this is the unfolding problem). Such spatial maps can help us better understand the structure of certain types of decision-making metrics employed in a variety of settings. This information can then be used to predict the outcome of future choices (for example, what product a certain consumer will buy or what candidate a voter will support) as an individual will be more likely to choose a stimulus that is closer to her position on a given dimension than one that is further away.

In Political Science MDS has a natural connection to spatial (geometric) models of choice and judgment. In the basic spatial model individuals/decision makers are assumed to have single-peaked utility functions over some latent dimensions of judgment and they choose the stimuli closest to them on the dimension(s) (Enelow and Hinich, 1984). For example, the methods developed in the past thirty years to estimate ideological positions from roll call data are examples of multidimensional unfolding applied to explicitly stated spatial models of choice (e.g., Poole and Rosenthal, 1997; Londregan, 2000; Clinton, Jackman and Rivers, 2004; Martin and Quinn, 2002; Pope and Treier, 2011).

In contrast to these modern methods that analyze nominal (Yea or Nay) data, the first uses of MDS methods in Political Science were to ratio scale data. For example, Weisberg and Rusk (1970) used non-metric MDS (Shepard, 1962a,b; Kruskal, 1964a,b) to analyze a correlation matrix computed over respondents' placements of political stimuli on 100 point "feeling" thermometers. Later Rabinowitz (1976), Cahoon, Hinich, and Ordeshook (1978), Jacoby (1982), and Brady (1990) developed unfolding methods based on the spatial theory of choice to analyze feeling thermometers.

We revisit these problems and show how to apply Bayesian methods to noisy ratio scale distances for both the similarities and unfolding problems. Our approach is explicitly based on the

spatial theory of choice and judgment. Because these problems have been studied for 50 years or more, the solutions are known and various data sets have been used to calibrate a succession of statistical methods. Both similarities and unfolding can easily be handled with frequentist or Bayesian models but pose problems for Markov chain Monte Carlo (MCMC) methods because of the existence of multiple posterior distributions that are mirror images of one another over the hyperplane of the parameters. In the on-line appendix we prove two theorems and two corollaries concerning the identification of posterior distributions in distance based choice models that allow us to isolate a single posterior distribution.

The analysis of ratio scale similarities data by psychometricians in the 1930s through the 1960s led to the development of multidimensional scaling methods (MDS). The psychometricians solved the general problem of representing relational or distance data in a spatial or geometric map where the points represented the stimuli and the distances between the points in the geometric map reproduced the observed distance/relational data. The ratio scale similarities problem was solved by Torgerson (1952, 1958) which in turn built upon work done by pschometricians in the 1930s [Eckart and Young, (1936); Young and Householder (1938)].

In the unfolding problem there are two sets of points – one representing individuals and one representing stimuli. The observed distance/relational data are regarded as expressing the preferences of individuals; namely, the closer a stimulus point is to an individual point the more the individual prefers that stimulus. The unfolding problem for ratio scale data (the "metric unfolding problem") was first solved by Schönemann (1970).

We first discuss the similarities problem and then we turn to the unfolding problem. Because our Bayesian framework is essentially the same for both problems, we spend more time detailing our solution for the similarities problem because the mathematical exposition is simpler. However, the unfolding problem is of greater interest because most public opinion survey data sets include a set of relational data questions in some form ("where would you place George Bush"; "On a scale of zero to 10, how would you rate John Kerry?"; etc.).

2. Multidimensional Bayesian Similarities/Dissimilarites Scaling

Similarities data differ from choice data in that they measure how alike or not-alike objects are to each other. Similarities data are frequently generated from choice data (for

example, an agreement score matrix based on how often legislators vote together) but does not include the choices themselves.

These data are relational and organized as a square matrix. For example, imagine you've never seen a map of the United States but you know the distances in miles between a set of U.S. cities. Within these distances is embedded a map and the purpose of multidimensional scaling (MDS) methods is producing a map given the distances. That is, in this instance, placing a set of points representing the cities in a plane such that they reproduce the distances. In this case MDS produces a geographical map that represents the distances in a meaningful way (i.e. the map would be useful for getting from one city to another). Clearly, a two dimensional map of distances between cities would be a simplification of the true positions of the cities as we exist in a three-dimensional world, but it is good enough for driving from one city to the next.

Table 1 displays distances in miles between five major U.S. cities. With just these data, we can generate a map using MDS techniques. However, there is an additional complication inherent in MDS techniques. The cities in Table 1 can be arranged in any number of different configurations that accurately represent the distances between them. Indeed, there

are an infinite number of arrangements of the points that reproduce the distances. As we describe below, this issue requires that we impose some constraints in order to produce a single, meaningful graphical representation and it also has additional implications for the Bayesian approach we present. In Figure 1 we plot four solutions from an MDS routine applied to the city distance data.

[TABLE 1 and FIGURE 1 HERE]

To anyone with cursory knowledge of U.S. geography, solution 4 is obviously the true configuration. Clearly there is no need to estimate the positions of the U.S. cities based on their distances as we know the exact positions of their coordinates (i.e. we know their latitude and longitude). Most applications of MDS to social science data, however, involve estimating positions on dimensions that are subjective in nature, for example, left-right ideological positions of legislators. As there is no 'true' set of coordinates in such settings, the relative placement of the points (as illustrated in figure 1) will always be arbitrary. In addition, the observed data will be noisy. Consequently, given a relative placement, it is highly desirable that the uncertainty in the distances is reflected in the estimated coordinates. In other words, we are not only interested in the relative positions of

certain objects (i.e. legislators) but also how certain we are about these estimated positions. This is particularly the case when using MDS solutions for statistical inference.

In Psychology various methods of multidimensional scaling (MDS) have been developed during the past 60 years to analyze similarities data. MDS methods model these similarities as distances between points in a geometric space (usually simple Euclidean). These MDS programs were designed to produce a picture or spatial map that summarizes the data graphically. MDS analyses are almost always in three dimensions or less because the whole *raison d'être* of MDS methods is to produce a visual summary. In this sense, a matrix of similarities/dissimilarities data can be regarded as having a geometric map embedded within it.

The purpose of an MDS method is to recover this geometric map. This requires a set of assumptions about the dissimilarities that allow them to be represented as points in an Euclidean space. The simplest assumption is just to model the observed dissimilarities as noisy realizations of the underlying true distances. For example, denote the observed dissimilarity (distance) as d^*_{jm} where

$$d^*_{jm} = d_{jm} + \varepsilon_{jm} \qquad (1)$$

Where j and m are both indices for the stimuli; i.e., j=1,…,q; m=1,…,q. Let Z_{jk} be the j^{th} stimulus coordinate on the k^{th} dimension, k=1,…,s, where s is the number of dimensions. Let d_{jm} be the Euclidean distance between stimulus j and stimulus m in the s-dimensional space:

$$d_{jm} = \sqrt{\sum_{k=1}^{s} (Z_{jk} - Z_{mk})^2} \qquad (2)$$

and

$$\varepsilon_{jm} = d^*_{jm} - d_{jm} \sim N(0, \sigma^2) \qquad (3)$$

or

$$\varepsilon_{jm} \sim \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}\left(d^*_{jm} - \sqrt{\sum_{k=1}^{s}(Z_{jk} - Z_{mk})}\right)^2}$$

This produces the likelihood function:

$$\mathbf{L^*(Z_{jk}|D^*)} = \frac{1}{(2\pi\sigma^2)^{\frac{q(q-1)/2}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(d^*_{jm} - \sqrt{\sum_{k=1}^{s}(Z_{jk} - Z_{mk})^2}\right)^2} \qquad (4)$$

Where $\mathbf{D^*}$ is the q by q matrix of observed dissimilarities. Taking the log of the right hand side and dropping the unnecessary constants yields a standard squared error loss function:

$$\ell n\xi = -\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(d^*_{jm} - \sqrt{\sum_{k=1}^{s}\left(Z_{jk} - Z_{mk}\right)^2}\right)^2 = -\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(d^*_{jm} - d_{jm}\right)^2 \qquad (5)$$

In standard optimization methods the first derivatives of (5) are used to find maximum likelihood estimates of the points. However, these derivatives are problematic because they all contain the ratio $\frac{d^*_{jm}}{d_{jm}}$ which is undefined when $Z_j = Z_m$ so that $d_{jm} = 0$. In practice this is not a problem but it and the fact that distances cannot be negative means that the statistical properties are not clear and that the assumption about the error, equation (3), is dicey at best. Nevertheless, finding Z's that minimize (or maximize as in equation (5)) the squared error loss function is relatively easy.

We now turn to a more realistic model of the data. We assume that the observed distances, d^*_{jm}, are drawn from the log-normal distribution because distances are inherently positive:

$$\ln(d^*_{jm}) \sim N(\ln(d_{jm}), \sigma^2) \qquad (6)$$

That is

$$f(d^*_{jm}) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}} d^*_{jm}} e^{\left(-\frac{1}{2\sigma^2}\left(\ln(d^*_{jm}) - \ln(d_{jm})\right)^2\right)}$$

Other researchers have used the truncated normal (Oh and Raftery, 2001), the normal (Navarro and Lee, 2003), and the

normal with an exponential mean (Okada and Shigemasu, 2010). We prefer the log-normal because we think it is more realistic model of the noise process; viz., the smaller the observed distance the smaller the variance of that distance.

Our likelihood function is:

$$\mathbf{L}^{*}(Z_{jk}|\mathbf{D}^{*})=\frac{1}{(2\pi\sigma^{2})^{\frac{q(q-1)/2}{2}}}\left(\prod_{j=1}^{q-1}\prod_{m=j+1}^{q}\frac{1}{d^{*}_{jm}}\right)e^{-\frac{1}{2\sigma^{2}}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d^{*}_{jm}\right)-\ln\left(\sqrt{\sum_{k=1}^{s}\left(Z_{jk}-Z_{mk}\right)^{2}}\right)\right)^{2}} \quad (7)$$

To implement our Bayesian model we use simple normal prior distributions for the stimuli coordinates:

$$\xi(Z_{jk}) = \frac{1}{(2\pi\kappa^{2})^{\frac{1}{2}}}e^{-\frac{Z_{jk}^{2}}{2\kappa^{2}}} \quad (8)$$

and an uniform prior for the variance term:

$$\xi(\sigma^{2})=\frac{1}{c},\ 0<c<b \quad (9)$$

where, empirically, b is no greater than 2.

Hence, our posterior distribution is:

$$\xi(Z_{jk}|\mathbf{D}^{*}) \propto \prod_{j=1}^{q-1}\prod_{m=j+1}^{q}\left\{f_{jm}(Z_{jm}\,|\,d^{*}_{jm})\right\}\xi(Z_{11})\xi(Z_{12})\ldots\xi(Z_{1s})\xi(Z_{21})\ldots\xi(Z_{qs})\xi(\sigma^{2}) \quad (10)$$

Taking the log of the right hand side and dropping the unnecessary constants:

$$\ell n\xi \propto -\frac{q(q-1)/2}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^{*}\right)-\ln\left(\sqrt{\sum_{k=1}^{s}\left(Z_{jk}-Z_{mk}\right)^2}\right)\right)^2$$

$$-\frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z_{jk}^2\right)-\ln(c)=$$

$$-\frac{q(q-1)/2}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^{*}\right)-\ln\left(d_{jm}\right)\right)^2-\frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z_{jk}^2\right)-\ln(c) \quad (11)$$

We experimented with vague and informative priors and found that the solutions were essentially identical for large matrices. We discuss this in more detail below. In the on-line appendix we show the first and second derivatives for (11). In our estimation work we check the solutions with both numerical and analytical first and second derivatives.

To illustrate our approach to similarities scaling, we use agreement scores computed between members of the U.S. 90th (1967-68) Senate. We chose the 90th Senate because it is well known that voting was two-dimensional during this period (Poole and Rosenthal, 1997). Given q roll call votes, the agreement score is the number of times a pair of senators vote the same way (Yea, Yea or Nay, Nay) divided by the number of times that they both voted on the same roll calls and multiplied by 100. The agreement scores range from 0 to 100 with 100 indicating

identical voting records. Table 2 shows a few Senators and their agreement scores.

[TABLE 2 HERE]

We convert the agreement scores to distances by subtracting them from 100 and dividing by 50. This is a convenient normalization because the estimated coordinates are usually in the unit hypersphere. Note that we include President Lyndon Johnson in the matrix by using Congressional Quarterly's presidential support roll calls. That is, CQ indicates on a fair number of roll calls whether a Yea/Nay is a vote in favor of the President's position. Hence, the President can be treated as a Senator. He just does not vote as often.

In this example, counting President Johnson, q is equal to 102. Our q by q symmetric matrix of distances has q(q-1)/2 unique entries (we ignore the diagonal of zeroes). Suppose there is an exact solution; that is, a set of q points in s dimensions that exactly reproduces the distances. Clearly, given that we only observe the distances, it does not matter what origin, rotation around that origin, or sign flips on the dimensions we select as long as the configuration of points vis a vis one another is not altered.

With q points in s dimensions we have to solve for q*s coordinates. However, we can set any point to the origin –

(0,0,…,0) – so this leaves us with q*s – s= (q-1)*s parameters. To pin down the configuration we need to set the rotation. In general, in *s* dimensions, with one point set to the origin, a rigid rotation of the configuration is determined by s-1 angles from the origin. In addition, given s-1 fixed angles in the rotation matrix, if we have a solution that exactly reproduces the matrix of distances then there are an additional 2^s-1 solutions that exactly reproduce the distances. These additional solutions are simply sign flips on the dimensions. For example, with *s*=2, suppose that we have a solution such that it reproduces our matrix of distances. Then there are three more solutions that also exactly reproduce the distances.

To fix the origin we set Senator Hill (D-AL) at the origin and we fix President Johnson's second dimension coordinate at zero which has the effect of fixing the angle from the origin in the rotation matrix. We use the Nelder-Mead (1965) amoeba method and the Powell (1973) method to obtain 1001 solutions for our log-posterior, equation (11), from random starts. The best solution and its reflections are shown in Figure 2. The tokens in the plots indicate the political party of the member -- "D" for northern Democrat, "S" for southern Democrat, and "R" for Republican. We computed both numerical and analytical first and second derivatives for the optimal solution to check that the

Hessian was full rank (i.e., negative-definite; see the on-line appendix).

[FIGURE 2 HERE]

Of the 1001 solutions for the Bayes posterior, only 3 were the solution (and its reflections) shown in figure 2. The log-likelihood was about -3100.0. The value for σ^2 was 0.1104. The extreme non-linearity of the log-normal likelihood function meant that a large number of modes were found by the optimizers. Many of these were quite close together in terms of log-likelihood.

In two dimensions fixing the origin and one coordinate of another point at zero is enough to pin down four identical posteriors corresponding to the sign flips as shown in figure 2. This is enough so that an optimizer can find modes but not enough for the efficient use of MCMC methods because of the reflections. That is, each member of the chain is a configuration. If the chain explores the entire hyperplane of the parameters then the means of all the coordinates will be zero because of the symmetry of the modes. However, in practice we can let the chain wander through the ((q-1)*s)-1 dimensional hyperplane and post-process the results by flipping the signs of each configuration in the chain back to a target configuration. This approach is very similar to that advocated by Bradlow and

Schmittlein (2000), Oh and Raftery (2001), Hoff, Raftery, and Handcock (2002), and Gormley and Murphy (2006).

For small similarities problems we found that in addition to the origin and one fixed coordinate simply adding three sign constraints to the three fixed coordinates isolated a single posterior. That is, keep the three constraints used to find the modes and then restrict three coordinates to be positive/negative. This works well and it is easy to implement in WinBUGS by using the I(,0) or I(0,) operators.

For larger problems like the 90th Senate agreement scores we retain the origin and one fixed coordinate and then solve for the sign flips by computing simple correlations dimension by dimension between coordinates from each draw in the chain (a configuration of points) and the coordinates from the optimizer solution.

The left panel of figure 3 shows the results for the 90th Senate (we adjusted the coordinates to -1 to +1 for presentation purposes). We ran multiple chains using a slice sampler (Neal, 2003) for 110,000 iterations and treated the first 10,000 draws as burn-in. The configuration is the mean of draws 10,001 to 110,000 (all the chains produced essentially identical results). The configuration is similar to that shown in figure 2. The variance term is precisely estimated with a standard deviation of 0.0026. The standard deviations around the points range from

about 0.08 to 0.18 with the largest being 0.25. Additionally, we assessed convergence using the Geweke, Heidelberger-Welch and Raftery and Lewis diagnostics. According to these diagnostics, the posteriors for all parameters meet all criteria for convergence. Also we experimented with informative priors (κ^2=1) and vague priors (κ^2=100) and the results did not change.

The right panel of figure 3 compares our Bayesian results with the best available MDS procedure; SMACOF (scaling by maximizing a convex function) developed by Jan De Leeuw and his colleagues (1977, 1988; De Leeuw and Heiser, 1988; De Leeuw and Mair, 2009). SMACOF is a sophisticated scaling method that minimizes the sum of squared error as given in equation (5) above (technically, the negative of equation (5)). SMACOF can perform non-metric MDS (Shepard, 1962a,b; Kruskal, 1964a,b) and has a flexible structure so that missing data is easily handled (See the R implementation, De Leeuw and Mair, 2009). We also use it below in our unfolding examples.

Our Bayesian configuration is essentially identical to the SMACOF configuration despite the very different models of the data. The advantage of a Bayesian approach is that we are able to get measures of uncertainty for the points. SMACOF is a classical scaling procedure that does not produce standard errors for the points. However, note in the Bayesian model fixing three coordinates has the effect of "transmitting" the

uncertainty associated with those coordinates to other points. There is no solution for this. It is just inherent in the problem.

[FIGURE 3 HERE]

Our approach has the advantage of isolating one posterior distribution and then analyzing it with standard MCMC methods. Our approach is computationally simple and can be implemented in publicly available software such as WinBUGS and JAGS. In the on-line appendix we show our WinBUGS script for the 90th Senate. We used informed priors derived from the Nelder-Mead configuration to stabilize the sampler in WinBUGS.

We now turn to a discussion of how to apply our approach to the unfolding problem.

3. A Bayesian Multidimensional Unfolding Model

In the unfolding problem we have two sets of points – one for individuals and one for stimuli. We are given only the noisy ratio scale distances between the two sets and not the distances within each set. We focus here on a Bayesian unfolding model based on the spatial model of choice. We assume that the individuals have symmetric single-peaked utility functions. Hence, we use an *ideal point* model in contrast to a *vector* model of choice. In a vector model the underlying utility of an

individual is not single-peaked. Rather, the utility increases through the space of the stimuli along the individual's vector (direction) through the space similar to the one-dimensional IRT model. Much work has been done on Bayesian vector and mixed models of unfolding in the marketing literature (Park, DeSarbo, and Liechty, 2008; Fong, DeSarbo, Park, and Scott, 2010; DeSarbo, Park, and Rao, 2010).

Denote the observed distance as d_{ij}^{*} where

$$d_{ij}^{*} = d_{ij} + \varepsilon_{ij} \qquad (12)$$

Where n is the number of individuals, i=1,…,n, and X_{ik} is the i^{th} individual coordinate on the k^{th} dimension. As before let Z_{jk} be the j^{th} stimulus coordinate on the k^{th} dimension, k=1,…,s, where s is the number of dimensions. Let d_{ij} be the Euclidean distance between individual i and stimulus j in the s-dimensional space:

$$d_{ij} = \sqrt{\sum_{k=1}^{s} (X_{ik} - Z_{jk})^2} \qquad (13)$$

As before, we assume that the observed distances, d_{ij}^{*}, are drawn from the log-normal distribution:

$$\ln(d_{ij}^{*}) \sim N(\ln(d_{ij}), \sigma^2)$$

Which produces the likelihood function:

$$L^*(X_{ik}, Z_{jk} | D^*) = \frac{1}{(2\pi\sigma^2)^{\frac{nq}{2}}} \left(\prod_{i=1}^{n} \prod_{j=1}^{q} \frac{1}{d_{ij}^*} \right) e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{q} \left(\ln\left(d_{ij}^*\right) - \ln\left(\sqrt{\sum_{k=1}^{s} \left(X_{ik} - Z_{jk}\right)^2} \right) \right)^2} \quad (14)$$

We use simple normal prior distributions for the individual and stimuli coordinates:

$$\xi(X_{ik}) = \frac{1}{(2\pi\zeta^2)^{\frac{1}{2}}} e^{-\frac{X_{ik}^2}{2\zeta^2}}$$

$$\xi(Z_{jk}) = \frac{1}{(2\pi\kappa^2)^{\frac{1}{2}}} e^{-\frac{Z_{jk}^2}{2\kappa^2}}$$

and an uniform prior for the variance term:

$$\xi(\sigma^2) = \frac{1}{c}, \; 0 < c < b$$

where b, empirically, is no greater than 2.

Hence, our posterior distribution is:

$$\xi(X_{ik}, Z_{jk} | D^*) \propto \prod_{i=1}^{n} \prod_{j=1}^{q} \left\{ f_{ij}(X_{ik}, Z_{jk} | d_{ij}^*) \right\} \xi(X_{11}) \ldots \xi(X_{ns}) \xi(Z_{11}) \ldots \xi(Z_{qs}) \xi(\sigma^2) \quad (15)$$

Taking the log of the right hand side and dropping the unnecessary constants:

$$\ell n\xi \propto -\frac{nq}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(\ln\left(d_{ij}^{*}\right) - \ln\left(\sqrt{\sum_{k=1}^{s}\left(X_{ik} - Z_{jk}\right)^2}\right)\right)^2$$
$$-\frac{1}{2\zeta^2}\left(\sum_{i=1}^{n}\sum_{k=1}^{s}X_{ik}^2\right) - \frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z_{jk}^2\right) - \ln(c) =$$

$$-\frac{nq}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(\ln\left(d_{ij}^{*}\right) - \ln\left(d_{ij}\right)\right)^2 - \frac{1}{2\zeta^2}\left(\sum_{i=1}^{n}\sum_{k=1}^{s}X_{ik}^2\right) - \frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z_{jk}^2\right) - \ln(c) \quad (16)$$

We experimented with vague and informative priors for both the ζ and the κ and found that the results were essentially the same. In the on-line appendix we show the first and second derivatives for (16).

Our unfolding example is the classic 1968 National Election Study feeling thermometers. A feeling thermometer asks individuals to respond to a set of stimuli (political figures in this case) based on their subjective views of warmth towards them. The thermometer ranges from 0 to 100 degrees with 100 indicating warm and very favorable feeling, 50 indicating neutrality towards the political figure, and 0 indicating that the respondent feels cold and very unfavorable towards the political figure. The 1968 feeling thermometers have been analyzed by Weisberg and Rusk (1970), Wang, et al. (1975), Rabinowitz (1976), Cahoon, et al. (1978), Poole and Rosenthal (1984), and Brady (1990) with the main focus on modeling the latent dimensions underlying the thermometers as well as testing theories of spatial voting.

In the NES 1968 survey twelve political figures were included in the thermometer questions: George Wallace, Hubert Humphrey, Richard Nixon, Eugene McCarthy, Ronald Reagan, Nelson Rockefeller, President Johnson, George Romney, Robert Kennedy, Edmund Muskie, Spiro Agnew, and Curtis LeMay. There were 1,673 respondents and we included in our analysis the 1,392 respondents who rated at least five of the twelve political figures.

We perform our analysis in two dimensions because previous analyses using optimization methods almost all find two dimensions in the data. We think this is due to the idiosyncratic noise in the thermometers (see Abrajano and Poole, 2011, for a discussion) and valence effects (Adams, Merrill, and Grofman, 2005; Londregan, 2000; Merrill and Grofman, 1999). A second dimension is picking up some of these effects and "smoothing" out the first dimension. Modeling valence effects is difficult so we leave that for future work. In any event, our aim here is to show the advantages of our Bayesian approach. Namely, a properly designed Markov chain reveals much more information than simply the modes of a loss function.

Finding the optimal solution for the unfolding problem is challenging because of the number of parameters. We used Limited-memory BFGS which can handle very large problems and is computationally efficient (see Liu and Nocedal, 1989, for a

discussion). In two dimensions this required a search over a 2,805 dimensional hyperplane. We set George Romney at the origin and Eugene McCarthy's second dimension coordinate at zero to obtain identification. We checked the first derivatives (see the on-line appendix) for the starting configuration to be sure that our points were located on a mode of the loss function, equation (16).

Using the respondent and candidate coordinates as targets we were able to run a slice sampler on the 1968 data (Neal, 2003). Because the ratio of respondents to the candidates is so large, we found that the method that worked the best was to first draw the respondent coordinates and then the candidate coordinates. We kept Romney at the origin but we did not constrain any other points because we found that simply rotating the drawn configuration to the optimal configuration with Romney at the origin was simple and easy to implement (see the on-line appendix for further detail). We ran our chain to 110,000 draws with the first 10,000 as burn-in. Figure 4 shows the results.

[FIGURE 4 HERE]

The upper left panel of Figure 4 shows the respondents along with tokens indicating the locations of Humphrey, Nixon, and Wallace. The lower left panel shows the candidate configuration. We display those respondents who indicated that they voted for Humphrey, Nixon, or Wallace, as the tokens "h",

"n", or "w", respectively. Humphrey, Nixon, and Wallace are located near where their voters are concentrated.

The upper right and lower right panels of Figure 4 show the results for an unfolding using SMACOF. Only Humphrey is located within where his voters are concentrated. In addition, the candidate configuration from SMACOF is bunched up save for Wallace and his running mate LeMay who are pushed outward. These differences arise from the fact that the simple additive model (the true distance plus noise) as in equation (12) does not work well with the thermometer data. This is due to the fact that there are a lot of thermometer scores close to 100. The log-normal is a more realistic model of the data because the standard error gets smaller as the distances get smaller.

We also tested the MAP unfolding procedure developed by Hinich (2005). MAP uses a maximum likelihood factor analysis of the covariance matrix computed from the squared distances to recover the stimuli and then OLS using the stimuli to recover the respondent locations. Although MAP is computationally efficient, the configuration of respondents produced by MAP is not as clean as that produced by SMACOF. We consider SMACOF to be a better unfolding procedure. Hence, we show only comparisons with SMACOF.

Figures 5 and 6 show thermometer unfolding results for the 2000 and 2004 elections. Again the Bayesian unfolding results

place the major candidates among their supporters and in these two elections the results are largely one dimensional. The SMACOF results tend to spread the voters out more on the second dimension but the location of the major presidential candidates is reasonable.

[FIGURES 5 and 6 HERE]

The candidates in the Bayesian unfoldings are precisely estimated. For example, in 1968 the largest standard deviation was for George Wallace's second dimension coordinate at 0.11. Figure 7 shows the bivariate densities for the three major 1968 Presidential candidates. All are unimodal over the two dimensional space.

[FIGURE 7 HERE]

The respondents were less precisely estimated. For example, Figure 8 shows the distribution of the 100,000 draws for the coordinates of respondent number 2 in 1968. The 2nd respondent was a young white male Democrat. He did not like Wallace, LeMay, Agnew, and Reagan (15, 30, 30, 30) but he was a little warmer towards Nixon and Romney (40, 40). He was less than enthused with President Johnson and Hubert Humphrey (50 and 60) but he really liked Robert Kennedy, Nelson Rockefeller, and Eugene McCarthy (97, 97, 85). His preferences roughly line up left to right but not entirely. This is reflected in the bimodal distribution of the draws. Most of the draws are in the

lower mode around -0.7, -0.6 with a smaller number around -0.5, 0.1. This shows the advantage of a Bayesian approach. A mode finder (optimization method) will land on one of the two modes whereas a Markov chain “illuminates” the entire distribution.

[FIGURE 8 HERE]

4. Conclusion

In this paper we have shown how to apply Bayesian methods to noisy ratio scale distances for both the classical similarities problem as well as the unfolding problem. Our approach combines the advantages of traditional mode finders and Bayesian MCMC. We use the mode finders to give us a target that identifies (“freezes”) the posterior for the Markov chain generator.

Our unfolding example using the 1968 candidate thermometers shows the power of MCMC (made possible by the speed of modern computers) to illuminate complex distributions. Instead of modes with their associated standard deviations from the inverse Hessian, "painting" the entire posterior distribution allows us to show means and the complete distribution of the parameters.

Our results are preliminary. We deliberately kept our models simple because our aim was to revisit older problems using modern methods. We think the thermometers are an underutilized resource that potentially can reveal important

information about individuals' utilities for political figures. Our aim here was simply to show a basic method that can be used as a springboard to more complex analyses.

Appendix to

BAYESIAN METRIC MULTIDIMENSIONAL SCALING

This Appendix shows a set of proofs of the existence of full rank Hessians for metric MDS in section A1; WINBUGS code for the Bayesian dissimilarities scaling of the 90^{th} Senate disagreement scores in A2; and the first and second derivatives for the similarities and unfolding models corresponding to equations (11) and (16) respectively, in A3.

In section A1 we show how to identify solutions for metric MDS problems. By identification what we mean is estimating the smallest number of parameters such that the Hessian matrix corresponding to a solution is full rank. If too many parameters are estimated the Hessian is singular. If too few are estimated then the log-posterior is distorted and a sub-optimal solution will be the result.

Assume that our dissimilarities data are squared distances between pairs of stimuli (the analysis of unfolding data is essentially the same). Our q by q symmetric matrix of data has q(q-1)/2 unique entries (we ignore the diagonal of zeroes). Suppose there is an exact solution; that is, a set of q points in s dimensions that exactly reproduces the squared distances.

Clearly, given that we only observe the distances, it does not matter what origin and the rotation around that origin we select as long as the configuration of points vis a vis one another is not altered.

With q points in s dimensions we have to solve for q*s parameters. However, we can set any point to the origin – (0,0,…,0) – so this leaves us with q*s – s= (q-1)*s parameters. To pin down the configuration we need to set the rotation. In general a rotation matrix is determined by s-1 angles from the origin and sign flips on each dimension. For example, in two dimensions the general form of the rotation matrix is:

$$\Gamma = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \; 0 \le \theta \le 2\pi$$

However, note that given a *specific* θ we have *four* rotation matrices:

$$\Gamma_1 = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad \Gamma_2 = \begin{bmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \Gamma_3 = \begin{bmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{bmatrix} \quad \Gamma_4 = \begin{bmatrix} -\cos\theta & -\sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

Or

$$\tilde{\Gamma} = \Delta\Gamma \text{ where } \Delta = \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}$$

That is, given a specific θ, there are 2^s sign flips corresponding to the s columns of the rotation matrix. With *s*=2, suppose that we have a solution $\tilde{Z}$ such that it reproduces our matrix of squared distances, **D**. Then there are three more solutions corresponding to the above rotation matrices that also exactly reproduce **D**. In general, in *s* dimensions, with one point set to the origin with s-1 fixed angles in the rotation matrix, if we have a solution $\tilde{Z}$ that exactly reproduces the matrix of squared distances then there an additional 2^s-1 solutions that exactly reproduce **D**.

This identification problem is very similar to that discussed by Rivers (2003). He discusses the identification of the classical maximum likelihood factor analysis problem and shows the number of restrictions necessary to get identification (these include fixing the origin and sign flips). However, his main concern is the identification of the multidimensional IRT model where the data are indicators and he shows that fixing s+1 points (or s(s+1) parameters) fully identifies the model. Our result is different because we assume that we observe (noisy) ratio scale data. Identification is somewhat different in this setting.

With these preliminaries we turn to our existence proofs.

A1: Existence Proofs for the Hessian

We have a total of (s*q)+1 parameters – the q points plus σ^2 (κ^2 is a fixed constant) for the similarities problem. For the unfolding problem we have (s*(n+q))+1 parameters. Because only distances, the $\boldsymbol{d}^*_{jm}$ and the d_{jm}, are used in the log-posterior, we impose the constraints that $\ln(d^*_{jm}) > 0, \forall j \neq m$ (or $\ln(d^*_{ij}) > 0, \forall i, j$) and $\ln(d_{jm}) > 0, \forall j \neq m$ (or $\ln(d_{ij}) > 0, \forall i, j$) for our proofs below. If any $\boldsymbol{d}^*_{jm}$ or d_{jm} is equal to zero for $j \neq m$ then equation (11) is equal to -∞. As a practical matter, this is not a problem for the observed data, $\boldsymbol{d}^*_{jm}$, because it can be rescaled or the corresponding j^{th} column and j^{th} row can be dropped on the assumption that the underlying j^{th} and m^{th} stimuli are the same.

For the unfolding problem if a $\boldsymbol{d}^*_{ij}$ is zero then equation (16) is equal to -∞. Again, as a practical matter the offending $\boldsymbol{d}^*_{ij}$ can be rescaled (e.g., set to a small distance greater than zero) or treated as missing data.

In our proofs below we analyze only the similarities problem because the unfolding problem is a subset of the similarities problem albeit with missing data. That is, we could set $\boldsymbol{W} = \begin{bmatrix} \boldsymbol{Z} \\ \boldsymbol{X} \end{bmatrix}$ where **W** is a (q+n) by s matrix and all the proofs would hold using **W** instead of **Z**.

In our proofs we assume that all the points are *distinct*; that is,

Definition: A set of points is distinct if $\boldsymbol{d_{jm} > 0 \; \forall j \neq m}$, or equivalently, $\forall$ j,m=1,…,q, and j≠m, $Z_j \neq Z_m$.

In practice distinctness is not a serious problem because if two points were the same, that is, $\mathbf{Z_j = Z_m}$, then there is a "pinhole" that goes down to -∞ in the surface of equation (11). Such a "pinhole" cannot be a maximum in any event. We simply avoid the problem by always "moving around" them.

Let $\ell n\xi^*$ denote the right hand side of equation (11). For any configuration of points in s dimensions, $\boldsymbol{Z_1, Z_2, \ldots, Z_{q-1}, Z_q}$, there is a unique $\hat{\sigma}^2$which is simply the mean of the q(q-1)/2 squared differences between $\ln(d^*_{jm})$ and $\ln(d_{jm})$ (see equation (A3) and (A12) in Appendix A3). Hence we will ignore it in the notation below and simply assume that it is computed from the configuration; that is, $\hat{\sigma}^2$**(Z)** or $\hat{\sigma}^2$**(Z,X)**.

Given a configuration of points in s dimensions, there are an infinite number of configurations that produce the same $\ell n\xi^*$ by adding a constant and rotating the original configuration. Let **Ω** be the set

$$\Omega = \left\{ \mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_{q-1}, \mathbf{Z}_q, \boldsymbol{\alpha}, \Gamma \right\} \qquad \text{(A1)}$$

where **α** is an s-length vector of additive constants and **Γ** is an s by s rotation matrix. Let $\ell n\xi^*(\Omega)$ be the function value for the set **Ω.** This allows us to state a simple non-existence theorem for the Hessian.

Theorem 1: Given Ω such that all Z are *distinct*, then the Hessian for any Ω that maximizes $\ell n\xi^*$ will be singular.

Proof: Given that there are an infinite number of configurations of points, there are an infinite number of **Ω**. However, since every possible configuration is a member of some **Ω** we can compute all possible values of $\ell n\xi^*(\Omega)$. Hence, it must be the case that for some **Ω***, $\ell n\xi^*(\Omega^*) \geq \ell n\xi^*(\Omega) \forall \Omega \neq \Omega^*$. However, no member of **Ω*** can be an inflection point because there are an infinite number of configurations in **Ω*** within any arbitrary distance from any selected configuration. Therefore the Hessian is singular for all members of **Ω***. Q.E.D.

Note that because $\ell n\xi^*(\Omega^*)$ is the value for every element of **Ω*** then this results in a uniform distribution over a subspace of the real q*s hyperplane of the parameters (much like a "mesa" but infinitely long). The same is true for other **Ω ≠ Ω***.

Geometrically, there are an infinite number of stacked uniform "mesas" over the q*s hyperplane of the parameters with **Ω*** having the highest "altitude" $\ell n\xi^*\left(\Omega^*\right)$. If one point is set to the origin then we still have an infinite number of stacked uniform "mesas" but now the radius of each "mesa" is finite with a value of $\sqrt{\sum_{j=1}^{q}\sum_{k=1}^{s} Z_{jk}^2}$ where one of the $Z_j=0$.

We now show that with q distinct points and s(s+1)/2 hard constraints the Hessian is full rank. Without loss of generality, we can pick **α** and **Γ** so that the q by s coordinate matrix, $\tilde{Z}$, has the following form:

$$\tilde{Z}=\begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & \cdots & Z_{1,q-1} & Z_{1q} \\ Z_{21} & Z_{22} & Z_{23} & \cdots & Z_{2,q-1} & Z_{2q} \\ & & \vdots & & & \\ Z_{q-s+1,1} & Z_{q-s+1,2} & Z_{q-s+1,3} & \cdots & Z_{q-s+1,q-1} & 0 \\ Z_{q-s,1} & Z_{q-s,2} & Z_{q-s,3} & \cdots & 0 & 0 \\ & & \vdots & & & \\ Z_{q-2,1} & Z_{q-2,2} & 0 & \cdots & 0 & 0 \\ Z_{q-1,1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \qquad \text{(A2)}$$

That is, we set Z_q to the origin and then pick s-1 angles for **Γ** such that all but one of the coordinates for Z_{q-1} are equal to zero, all but two of the coordinates for Z_{q-2} are equal to zero, and so on. As we explained above, we have the sign flips,

$\tilde{\Gamma} = \Delta\Gamma$, where **Δ** is an s by s diagonal matrix with plus or minus ones on the diagonal. Given s-1 specific angles, θ_1 , θ_2 ,..., θ_{s-1} , then there are 2^s sign flips corresponding to the s columns of the rotation matrix. This allows us to define

$$\Omega = \left\{ \tilde{\boldsymbol{Z}}, \Delta \right\} \qquad \text{(A3)}$$

Theorem 2: Given Ω as in (A3) such that all $\tilde{Z}$ are *distinct*, then the Hessian for any Ω that maximizes $\ell n\xi^*$ will be rank q*s-(s*(s+1)/2).

Proof: Every configuration of points, **Z**, can be transformed into $\tilde{\boldsymbol{Z}}$ as in equation (A2) by choice of origin and rotation without changing the inter-point distances. Given that there are an infinite number of configurations of points, **Z**, and each one can be transformed into a $\tilde{\boldsymbol{Z}}$ that satisfy equation (A2), there are an infinite number of **Ω** that satisfy equation (A3). However, since every possible $\tilde{\boldsymbol{Z}}$ configuration is a member of some **Ω**, then we can compute all possible values of $\ell n\xi^*(\Omega)$. Hence, it must be the case that for some **Ω***, $\ell n\xi^*(\Omega^*) \geq \ell n\xi^*(\Omega) \forall \Omega \neq \Omega^*$. By construction, given s-1 specific angles, θ_1 , θ_2 ,..., θ_{s-1} , then there are 2^s sign flips corresponding to the s columns **Δ**. Hence, **Ω*** has 2^s members and each is separated from the others by a distance of at least

$\delta = \min\left\{2\sqrt{\sum_{j=1}^{q-1}\tilde{Z}_{j1}^2}, 2\sqrt{\sum_{j=1}^{q-1}\tilde{Z}_{j2}^2}, ..., 2\sqrt{\sum_{j=1}^{q-1}\tilde{Z}_{js}^2}\right\} > 0$. Denote these configurations as $\tilde{Z}^*$ so that $\Omega^* = \left\{\tilde{Z}^*, \Delta\right\}$. Consider any nearby configuration of distinct points, $\tilde{Z}$, $\Omega = \left\{\tilde{Z}, \Delta\right\}$, within an infinitesimal distance of **Ω***; that is, $\gamma = \sqrt{\sum_{j=1}^{q}\sum_{k=1}^{s}\left(\tilde{Z}_{jk}^* - \tilde{Z}_{jk}\right)^2} > 0$. Hence, by construction $\ell n\xi^*\left(\Omega^*\right) > \ell n\xi^*\left(\Omega\right)$. Because this is true for an infinitesimal distance on the q*s-(s*(s+1)/2) dimensional hyperplane in any direction from $\tilde{Z}^*$, the 2^S members of **Ω*** are inflection points with corresponding full rank Hessians. Q.E.D.

Note that the key difference between Theorems 1 and 2 is that **Ω*** in Theorem 1 had an infinite number of members and in Theorem 2 **Ω*** had 2^S members. In Theorem 1 this meant that no member of **Ω*** could be an inflection point because there are an infinite number of members within an infinitesimal distance of any selected member (the "mesa"). In contrast, the 2^S members of **Ω*** from Theorem 2 are separated from each other by non-zero distances. Hence, it is easy to show using a standard argument from mathematical analysis that any configuration not in **Ω*** that is an infinitesimal distance from one of the 2^S members of **Ω*** must be, by construction, less than the maximum; that is, $\ell n\xi^*\left(\Omega^*\right) > \ell n\xi^*\left(\Omega\right)$.

We now show two corollaries: first, if the number of hard constraints is less than s(s+1)/2, then the Hessian is singular; and second, if the number of hard constraints is greater than s(s+1)/2 then the solution is inferior in the sense that $\ell n\xi^*\left(\Omega^*\right) > \ell n\xi^*\left(\Omega\right)$.

Corollary 1: Let the number of hard constraints be less than s(s+1)/2. Given Ω such that all Z are *distinct*, then the Hessian for any Ω that maximizes $\ell n\xi^*$ will be singular.

Proof: Suppose that the number of hard constraints is (s(s+1)/2)-1. Without loss of generality modify $\tilde{\boldsymbol{Z}}$ so that $-\infty < Z_{q-s+1,q} < +\infty$, that is, coordinate $Z_{q-s+1,q}$ is not constrained to be zero. Denote this modified configuration as $\tilde{\boldsymbol{Z}}_{(-1)}$. There are an infinite number of Ω that satisfy $\Omega = \left\{\tilde{\boldsymbol{Z}}_{(-1)}, \Delta\right\}$. However, since every possible $\tilde{\boldsymbol{Z}}_{(-1)}$ configuration is a member of some Ω, then we can compute all possible values of $\ell n\xi^*\left(\Omega\right)$. Hence, it must be the case that for some **Ω***, $\ell n\xi^*\left(\Omega^*\right) \geq \ell n\xi^*\left(\Omega\right) \forall \Omega \neq \Omega^*$. **Ω*** has an infinite number of members because, by construction, $-\infty < Z_{q-s+1,q} < +\infty$. However, no member of **Ω*** can be an inflection point because there are an infinite number of configurations in **Ω*** within any arbitrary infinitesimal distance from any selected

configuration in the direction of $Z_{q-s+1,q}$. Therefore the Hessian is singular for all members of **Ω***. Finally, it is easy to construct similar arguments for $\tilde{\boldsymbol{Z}}_{(-2)}$, $\tilde{\boldsymbol{Z}}_{(-3)}$, etc. Q.E.D.

Corollary 2: Let the number of hard constraints be greater than s(s+1)/2. Given Ω such that all Z are *distinct*, then the Ω that maximizes $\ell n\xi^*$ will be less than $\ell n\xi^*$ for a Z with (s+1)/2 hard constraints as in equation (A2).

Proof: Suppose that the number of hard constraints is (s(s+1)/2)+1. Without loss of generality modify $\tilde{\boldsymbol{Z}}$ so that an additional coordinate is constrained to be a constant; for example, let $\boldsymbol{Z}_{\boldsymbol{q-s+1,q-1}} = \boldsymbol{C}_{\boldsymbol{q-s+1,q-1}}$. Denote this modified configuration as $\tilde{\boldsymbol{Z}}_{(+1)}$. However this modified configuration is a member of some Ω used in Theorem 2. From Theorem 2 we have $\ell n\xi^*\left(\Omega^*\right) \geq \ell n\xi^*\left(\Omega\right) \forall \Omega \neq \Omega^*$. Hence, unless $\boldsymbol{C}_{\boldsymbol{q-s+1,q-1}}$ is exactly equal to $\boldsymbol{Z}_{\boldsymbol{q-s+1,q-1}}$ in **Ω*** $\ell n\xi^*\left(\Omega^*\right) > \ell n\xi^*\left(\Omega\right)$. Finally, it is easy to construct similar arguments for any subset of additionally constrained coordinates. Q.E.D.

In one dimension, setting one point to the origin results in two solutions with the same $\ell n\xi^*$ values. For purposes of characterizing the distributions of the parameters with MCMC

methods, setting the sign -- a soft constraint -- on a second point (typically a point that is distant from the origin) isolates one log-posterior. So we get a unique log-posterior with one hard and one soft constraint. However, note that, if we use two hard constraints by fixing two points we get an inferior result because we have fixed one of the distances.

A2: WINBUGS SIMILARITIES MODEL

```
#
#  MDS Model for 90th Senate--over constrained
#
model{

#  Fix one point
#
        x[8,1] <- -0.626000480
……………  x[8,2] <-  0.46524749
#
# llh and sumllh monitor the log-likelihood
#
for (i in 1:101){
    llh[i,i] <- 0.0
    for (j in i+1:102){
#
#  Read in Distances rather than the similarities (makes handling missing data easier)
#
         dstar[i,j] ~ dlnorm(mu[i,j],tau)
         mu[i,j] <- log(sqrt((x[i,1]-x[j,1])*(x[i,1]-x[j,1])+(x[i,2]-x[j,2])*(x[i,2]-x[j,2])))
         llh[i,j] <- (log(dstar[i,j])-mu[i,j])*(log(dstar[i,j])-mu[i,j])
         llh[j,i] <- (log(dstar[i,j])-mu[i,j])*(log(dstar[i,j])-mu[i,j])
    }
}

   llh[102,102] <- 0.0
   sumllh <- sum(llh[,])
#
  ## priors
  tau ~ dgamma(1,1)

#
# Informed priors placed below (not all shown)
#
  x[1,1] ~ dnorm(0,.1) I(0,)
  x[1,2] ~ dnorm(0,.1) I(,0)
  x[2,1] ~ dnorm(0,.1) I(,0)
  x[2,2] ~ dnorm(0,.1) I(0,)

...etc. etc.

  x[98,1] ~ dnorm(0,.1) I(,0)
  x[98,2] ~ dnorm(0,.1) I(0,)
  x[99,1] ~ dnorm(0,.1) I(,0)
  x[99,2] ~ dnorm(0,.1) I(, -0.5)
  x[100,1] ~ dnorm(0,.1) I(,0)
  x[100,2] ~ dnorm(0,.1) I(,-0.5)
  x[101,1] ~ dnorm(0,.1) I(0.5,)
  x[101,2] ~ dnorm(0,.1) I(0.2,)
  x[102,1] ~ dnorm(0,.1) I(,-0.2)
  x[102,2] ~ dnorm(0,.1) I(,0)

}
```

A3: The Derivatives for the Log-Normal Bayesian Model

Similarities: The first derivatives for the similarities problem are:

$$\frac{\partial \ell n\xi}{\partial Z_{jk}} = -2\frac{1}{2\sigma^2}\sum_{j\neq m}^{q}\left\{\left(\ln\left(d_{jm}^{*}\right)-\ln\left(d_{jm}\right)\right)\left(-\frac{1}{d_{jm}}\right)\left(\frac{1}{2}\right)\left[\sum_{k=1}^{s}(Z_{jk}\text{ - }Z_{mk})^2\right]^{-\frac{1}{2}}\left(2\left[Z_{jk}\text{ - }Z_{mk}\right]\right)\right\}-\frac{Z_{jk}}{\kappa^2}$$

which simplifies to

$$\frac{\partial \ell n\xi}{\partial Z_{jk}} = \frac{1}{\sigma^2}\sum_{j\neq m}^{q}\left\{\frac{\left(\ln\left(d_{jm}^{*}\right)-\ln\left(d_{jm}\right)\right)}{d_{jm}^{2}}\left(Z_{jk}\text{ - }Z_{mk}\right)\right\}-\frac{Z_{jk}}{\kappa^2} \qquad \textbf{(A1)}$$

and

$$\frac{\partial \ell n\xi}{\partial \sigma^2} = -\frac{q(q-1)}{4\sigma^2}+\frac{1}{2\sigma^4}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^{*}\right)-\ln\left(d_{jm}\right)\right)^2 \qquad \textbf{(A2)}$$

Hence, we get the usual result for the variance term:

$$\hat{\sigma}^2 = \frac{2}{q(q-1)}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(ln\left(d_{jm}^{*}\right)-ln\left(d_{jm}\right)\right)^2 \qquad \textbf{(A3)}$$

Note that if κ^2 is a vague prior, the practical effect is that at an inflection point we have $\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\sigma^2} \approx \frac{\partial \ell n\xi}{\partial Z_{jk}} = 0$. Numerically, this is a handy result because it makes computing the inverse Hessian much easier to accomplish.

The second derivative for the variance is:

$$\frac{\partial^2 \ell n\xi}{\partial\sigma^2\partial\sigma^2} = \frac{q(q-1)}{4\sigma^4} - \frac{1}{\sigma^6}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)^2 \qquad \textbf{(A4)}$$

Substituting (A3) into (A4) it is easy to show that $\frac{\partial^2 \ell n\xi}{\partial\sigma^2\partial\sigma^2} < 0$ so that when the global maximum for the Z_{jk} is found σ^2 will be a maximum as well.

The second derivatives for the coordinates are:

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{jk}} = -4\sum_{j\neq m}^{q}\frac{\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)}{d_{jm}^4}\left(Z_{jk} - Z_{mk}\right)^2 - 2\sum_{j\neq m}^{q}\left[\frac{\left(Z_{jk} - Z_{mk}\right)^2}{d_{jm}^4}\right] + 2\sum_{j\neq m}^{q}\left[\frac{\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)}{d_{jm}^2}\right] - \frac{1}{\kappa^2}$$

(A5)

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{mk}} = 4\frac{\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)}{d_{jm}^4}\left(Z_{jk} - Z_{mk}\right)^2 + 2\frac{\left(Z_{jk} - Z_{mk}\right)^2}{d_{jm}^4} - 2\frac{\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right)}{d_{jm}^2} \qquad \textbf{(A6)}$$

In more than one dimension

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{j\ell}} = -2\sum_{j\neq m}^{q}\left\{\left[\frac{\left(Z_{jk} - Z_{mk}\right)\left(Z_{j\ell} - Z_{m\ell}\right)}{d_{jm}^4}\right]\left[2\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right) - 1\right]\right\} \qquad \textbf{(A7)}$$

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{m\ell}} = 2\frac{\left(Z_{jk} - Z_{mk}\right)\left(Z_{j\ell} - Z_{m\ell}\right)}{d_{jm}^4}\left[2\left(\ln\left(d_{jm}^*\right) - \ln\left(d_{jm}\right)\right) + 1\right] \qquad \textbf{(A8)}$$

where $\ell = 1,\ldots,s$ and $\ell \neq k$.

Unfolding: The first derivatives for the unfolding problem are:

$$\frac{\partial \ell n\xi}{\partial X_{ik}} = \frac{1}{\sigma^2}\sum_{j=1}^{q}\left\{\frac{\left(\ln\left(d_{ij}^*\right) - \ln\left(d_{ij}\right)\right)}{d_{ij}^2}\left(X_{ik} - Z_{jk}\right)\right\} - \frac{X_{ik}}{\zeta^2} \qquad \textbf{(A9)}$$

$$\frac{\partial \ell n\xi}{\partial Z_{jk}} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\left\{\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{2}}\left(X_{ik} - Z_{jk}\right)\right\} - \frac{Z_{jk}}{\kappa^2} \qquad \textbf{(A10)}$$

and

$$\frac{\partial \ell n\xi}{\partial \sigma^2} = -\frac{nq}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)^2 \qquad \textbf{(A11)}$$

Hence, we get the usual result for the variance term for the unfolding model:

$$\hat{\sigma}^2 = \frac{1}{nq}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(ln\left(d_{ij}^{*}\right)-ln\left(d_{ij}\right)\right)^2 \qquad \textbf{(A12)}$$

Note that if ζ^2 and κ^2 are vague priors, the practical effect is that at an inflection point we have $\frac{\partial^2 \ell n\xi}{\partial X_{ik}\sigma^2} \approx \frac{\partial \ell n\xi}{\partial X_{ik}} = 0$.

The second derivative for the variance is:

$$\frac{\partial^2 \ell n\xi}{\partial \sigma^2 \partial \sigma^2} = \frac{nq}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}\sum_{j=1}^{q}\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)^2 \qquad \textbf{(A13)}$$

Substituting (A12) into (A13) it is easy to show that $\frac{\partial^2 \ell n\xi}{\partial \sigma^2 \partial \sigma^2} < 0$ so that when the global maximum for the X_{ik} and Z_{jk} is found σ^2 will be a maximum as well.

The second derivatives for the coordinates are:

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial X_{ik}} = -2\sum_{j=1}^{q}\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{4}}\left(X_{ik}-Z_{jk}\right)^{2} - \sum_{j=1}^{q}\left[\frac{\left(X_{ik}-Z_{jk}\right)^{2}}{d_{ij}^{4}}\right] + \sum_{j=1}^{q}\left[\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{2}}\right] - \frac{1}{\zeta^{2}} \quad \textbf{(A14)}$$

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{jk}} = 2\sum_{i=1}^{n}\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{4}}\left(X_{ik}-Z_{jk}\right)^{2} + \sum_{i=1}^{n}\left[\frac{\left(X_{ik}-Z_{jk}\right)^{2}}{d_{ij}^{4}}\right] - \sum_{i=1}^{n}\left[\frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{2}}\right] - \frac{1}{\kappa^{2}} \quad \textbf{(A15)}$$

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial Z_{jk}} = \left[\frac{\left(X_{ik}-Z_{jk}\right)^{2}}{d_{ij}^{4}}\right]\left[2\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)+1\right] - \frac{\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)}{d_{ij}^{2}} \quad \textbf{(A16)}$$

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial X_{hk}} = \frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{mk}} = 0 \quad \textbf{(A17)}$$

Where h=1,...,n and h≠i. In more than one dimension

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial X_{i\ell}} = -\sum_{j=1}^{q}\left\{\left[\frac{\left(X_{ik}-Z_{jk}\right)\left(X_{i\ell}-Z_{j\ell}\right)}{d_{ij}^{4}}\right]\left[2\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)+1\right]\right\} \quad \textbf{(A18)}$$

$$\frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{j\ell}} = -\sum_{i=1}^{n}\left\{\left[\frac{\left(X_{ik}-Z_{jk}\right)\left(X_{i\ell}-Z_{j\ell}\right)}{d_{ij}^{4}}\right]\left[2\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)+1\right]\right\} \quad \textbf{(A19)}$$

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial Z_{j\ell}} = \left[\frac{\left(X_{ik}-Z_{jk}\right)\left(X_{i\ell}-Z_{j\ell}\right)}{d_{ij}^{4}}\right]\left[2\left(\ln\left(d_{ij}^{*}\right)-\ln\left(d_{ij}\right)\right)+1\right] \quad \textbf{(A20)}$$

$$\frac{\partial^2 \ell n\xi}{\partial X_{ik}\partial X_{h\ell}} = \frac{\partial^2 \ell n\xi}{\partial Z_{jk}\partial Z_{m\ell}} = 0 \quad \textbf{(A21)}$$

APPENDIX REFERENCES

Rivers, Douglas. 2003. "Identification of Multidimensional Spatial Voting Models." Working Paper, Stanford University (http://polmeth.wustl.edu/media/Paper/river03.pdf).

References

Abrajano, Marisa and Keith T. Poole. 2011. "A Method of Linking Surveys Using Affective Signatures with an Application to Racial/Ethnic Groups in the U.S." In *Who Gets Represented?* Edited by Peter Enns and Christopher Wlezien. New York: Russell Sage Foundation

Adams, James, Samuel Merrill and Bernard Grofman. 2005. *A Unified Theory of Party Competition: A Cross-National Analysis Integrating Spatial and Behavioral Factors*. New York: Cambridge University Press.

Bakker, Ryan and Keith T. Poole. 2012. "Replications Materials: (also available at http://voteview.com/Bakker_Poole_Bayesian_MDS.htm).

Bradlow, Eric T. and David C. Schmittlein. 2000. "The Little Engines that Could: Modeling the Performance of World Wide Web Search Engines." *Marketing Science*, 19(1):43-62.

Brady, Henry. 1990. "Traits Versus Issues: Factor Versus Ideal-Point Analysis of Candidate Thermometer Ratings." *Political Analysis*, 2:97-129.

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*, ed. by P.C. Wang. New York, Academic Press.

Clinton, Joshua D., Simon D. Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data: A Unified Approach." *American Political Science Review,* 98:355-370.

De Leeuw, Jan. 1977. "Applications of Convex Analysis to Multidimensional Scaling." In J.R. Barra, F. Brodeau, G. Romier, and B. van Cutsem (Eds.), *Recent Developments in Statistics* (pp. 133-145). Amsterdam, The Netherlands: North-Holland.

De Leeuw, Jan. 1988. “Convergence of the Majorization Method for Multidimensional Scaling.” *Journal of Classification,* 5:163-180.

De Leeuw, Jan and Willem Heiser. 1977. “Convergence of Correction-Matrix Algorithms for Multidimensional Scaling.” In J.C. Lingoes, E.E. Roskam, and I. Borg (Eds.), *Geometric Representations of Relational Data* (pp. 735-752). Ann Arbor, MI: Mathesis Press.

de Leeuw, Jan & Mair, P. 2009. Multidimensional scaling using majorization: The R package smacof. Journal of Statistical Software, 31(3), 1-30, http://www.jstatsoft.org/v31/i03/

DeSarbo, Wayne S., Joonwook Park, and Vithala R. Rao. 2011. “Deriving Joint Space Positioning Maps from Consumer Preference Ratings.” *Marketing Letters*, 22(1):1-14.

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

Enelow, James M. and Melvin J. Hinich. 1984. *The Spatial Theory of Voting.* New York: Cambridge University Press.

Fong, Duncan K. H., Wayne S. DeSarbo, Joonwook Park, and Crystal J. Scott. 2010. “A Bayesian Vector Multidimensional Scaling Procedure for the Analysis of Ordered Preference Data.” *Journal of the American Statistical Association,* 105(490):482-492.

Gormley, Isobel C. and Thomas B. Murphy. 2006. “A Latent Space Model for Rank Data.” *Statistical Network Analysis: Models, Issues and New Directions. Lecture Notes in Computer Science*. New York: Springer-Verlag.

Hinich, Melvin J. 2005. “A New Method for Statistical Multidimensional Unfolding.” *Communications in Statistics—Theory and Methods,* 34:2299-2310.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. “Latent Space Approaches to Social Network Analysis.” *Journal of the American Statistical Association*, 97 (No. 460):1090-1098.

Jacoby, William G. 1982. “Unfolding the Party Identification Scale: Improving the Measurement an Important Concept.” *Political Methodology,* 8:33-60.

Kruskal, Joseph B. 1964a. “Multidimensional Scaling by Optimizing a Goodness of Fit to a Nonmetric Hypothesis.” *Psychometrika*, 29:1-27.

Kruskal, Joseph B. 1964b. “Nonmetric Multidimensional Scaling: A Numerical Method.” *Psychometrika*, 29:115-129.

Liu, Dong C. and Jorge Nocedal. 1989. “On the Limited Memory BFGS Method.” *Mathematical Programming*, 45(1-3):503-528.

Londregan, John B. 2000. "Estimating Legislators' Preferred Points." *Political Analysis*, 8(1):35-56.

Martin, Andrew D. and Kevin M. Quinn. 2002. “Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999.” *Political Analysis*, 10:134-153.

Merrill, Samuel III and Bernard Grofman. 1999. *A Unified Theory of Voting: Directional and Proximity Spatial Models*. New York: Cambridge University Press.

Navarro, Daniel J. and Michael D. Lee. 2003. “Combining Dimensions and Features in Similarity-Based Representations.” In S. Becker, S. Thrun, and K. Obermayer (Eds.), *Advances in Neural Information Processing Systems,* 15:59-66. Cambridge, MA: MIT Press.

Neal, Radford M. 2003. “Slice Sampling.” *The Annals of Statistics* 31(3): 705-767.

Nelder, John A. and Roger Mead. 1965. "A simplex method for function minimization." *Computer Journal* **7**: 308–313.

Oh, Man-Suk and Adrian E. Raftery. 2001. “Bayesian Multidimensional Scaling and Choice of Dimension.” *Journal of the American Statistical Association*, 96 (No. 455): 1031-1044.

Okada, Kensuke and Kazuo Shigemasu. 2010. “Bayesian Multidimensional Scaling for the Estimation of a Minkowski Exponent.” *Behavior Research Methods*, 42(4):899-905.

Park, Joonwook, Wayne S. DeSarbo, and John Liechty. 2008. “A Hierarchical Bayesian Multidimensional Scaling Methodology for Accommodating both Structural and Preference Heterogeneity.” *Psychometrika*, 73(3):451-472.

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.

Pope, Jeremy C. and Shawn A. Treier. 2011. “Reconsidering the Great Compromise at the Federal Convention of 1787: Deliberation and Agenda Effects on the Senate and Slavery.” *American Journal of Political Science*, 55:289-306.

Powell, Michael J. D. 1973. ”On Search Directions for Minimization Algorithms.” *Mathematical Programming* 4: 193–201.

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

Schönemann, Peter H. 1970. “On Metric Multidimensional Unfolding.” *Psychometrika,* 35:349-366.

Shepard, Roger N. 1962a. “The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function. I.” *Psychometrika*, 27:125-139.

Shepard, Roger N. 1962b. “The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function. II.” *Psychometrika*, 27: 219-246.

Torgerson, Warren S. 1952. “Multidimensional Scaling: I. Theory and Method. *Psychometrika,* 17:401-419.

Torgerson, Warren S. 1958. *Theory and Methods of Scaling*. New York: Wiley.

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 Alston S. Householder. 1938. “Discussion of a Set of Points in Terms of their Mutual Distances.” *Psychometrika*, 3:19-22.

Tables

Table 1: Driving Distances Between 5 Cities

City	Boston	Detroit	Chicago	SF	Miami
Boston	0	702	983	3179	1539
Detroit	702	0	279	2475	1409
Chicago	983	279	0	2212	1309
SF	3179	2475	2212	0	3097
Miami	1539	1409	1390	3097	0

Table 2: Agreement Scores for 90th Senate (Partial)

JOHNSON (D-Pres)	100	61	50	52	65	70	37	...
SPARKMAN (D-AL)	61	100	89	50	65	85	65	...
HILL (D-AL)	50	89	100	53	62	78	69	...
GRUENING (D-AK)	52	50	53	100	76	58	43	...
BARTLETT (D-AK)	65	65	62	76	100	70	47	...
HAYDEN (D-AZ)	70	85	78	58	70	100	57	...
FANNIN (R-AZ)	37	65	69	43	47	57	100	...

FIGURES

Figure 1: Solutions for 5 Cities Data

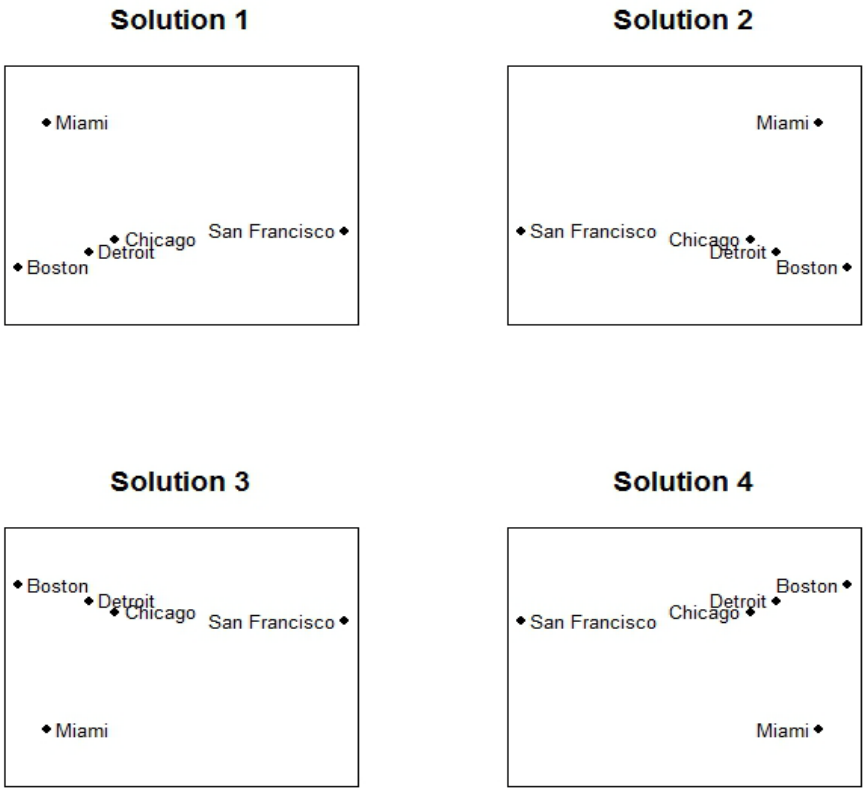

Figure 2: Best 90th Senate Configuration and its Reflections

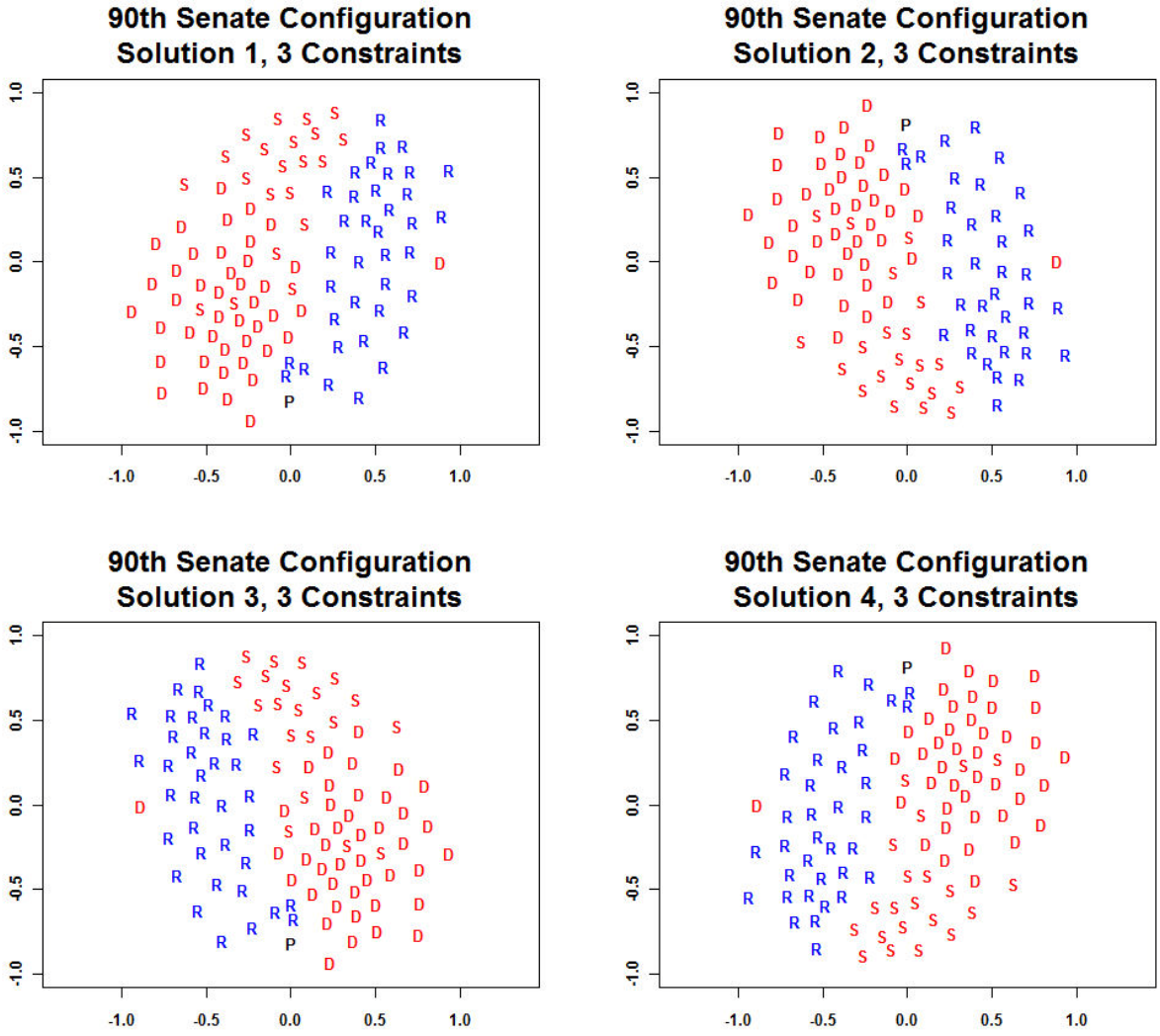

Figure 3: 90th Senate Bayesian and SMACOF Solutions

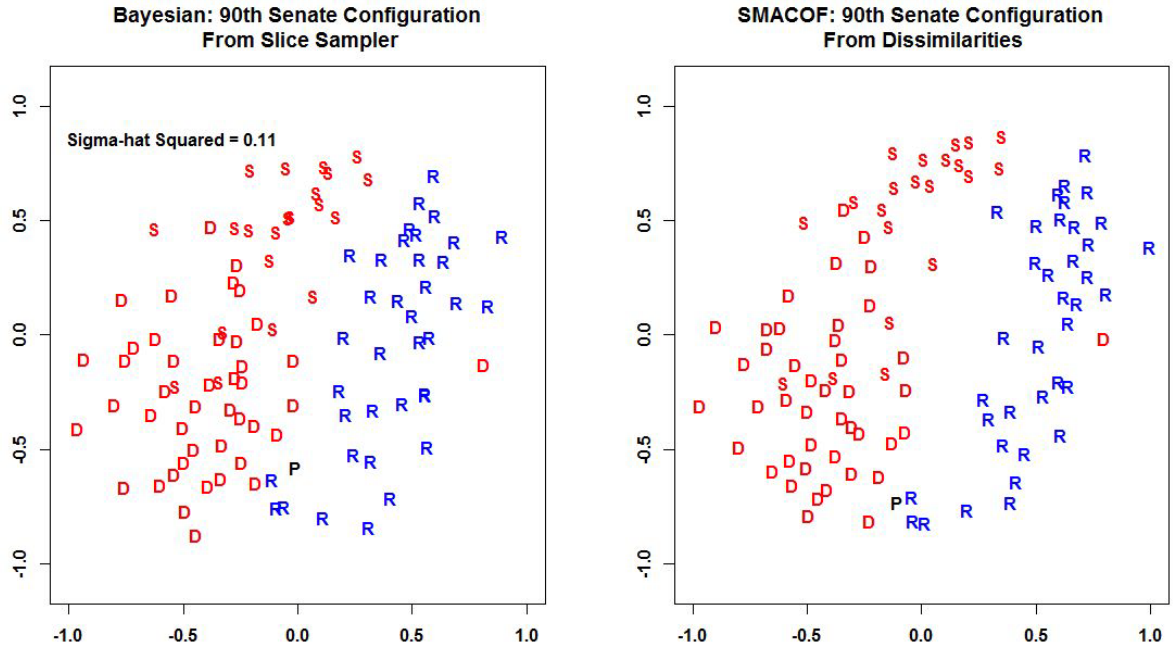

Figure 4: 1968 Thermometer Bayesian vs. SMACOF Unfolding

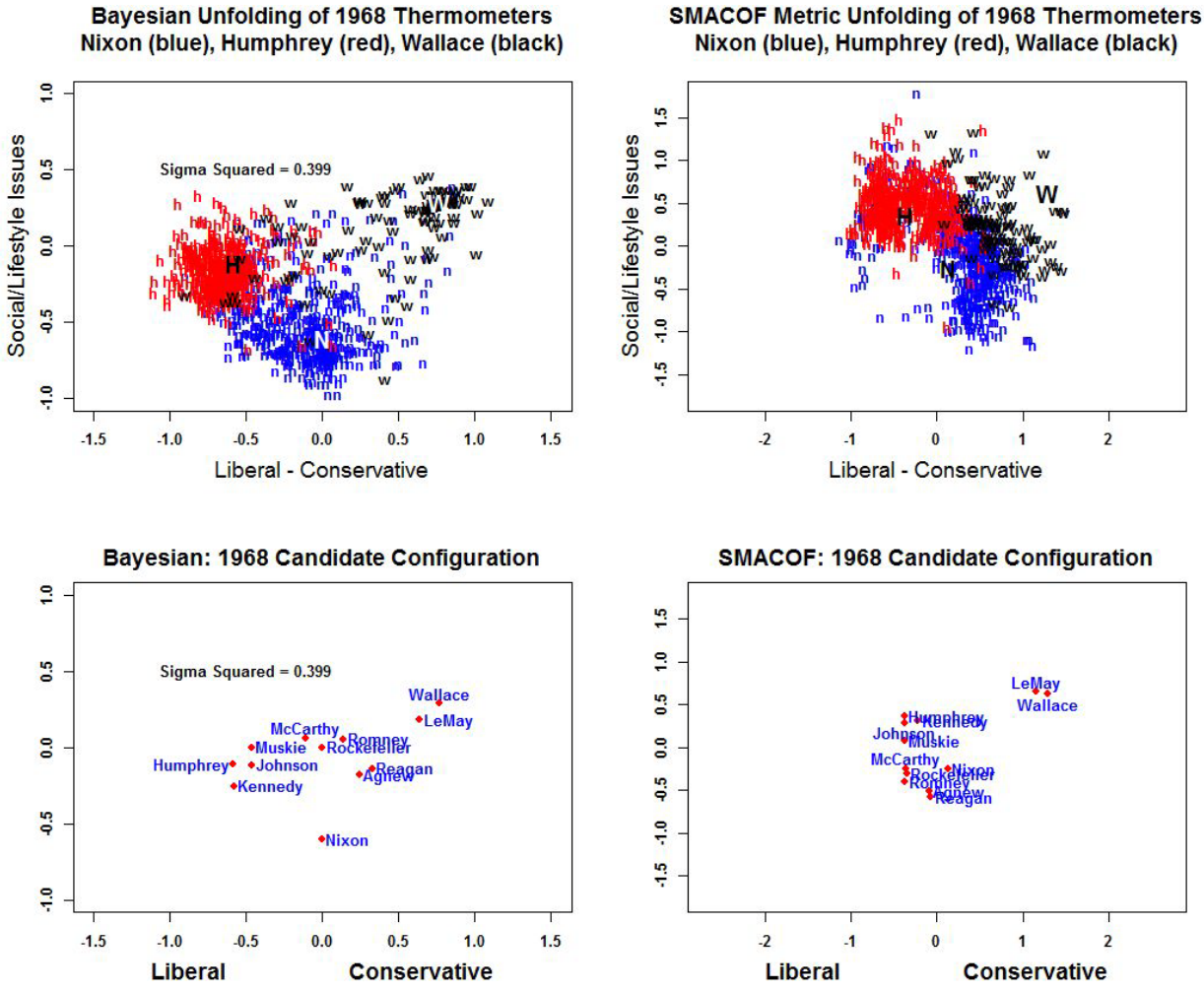

Figure 5: 2000 Thermometer Bayesian vs. SMACOF Unfolding

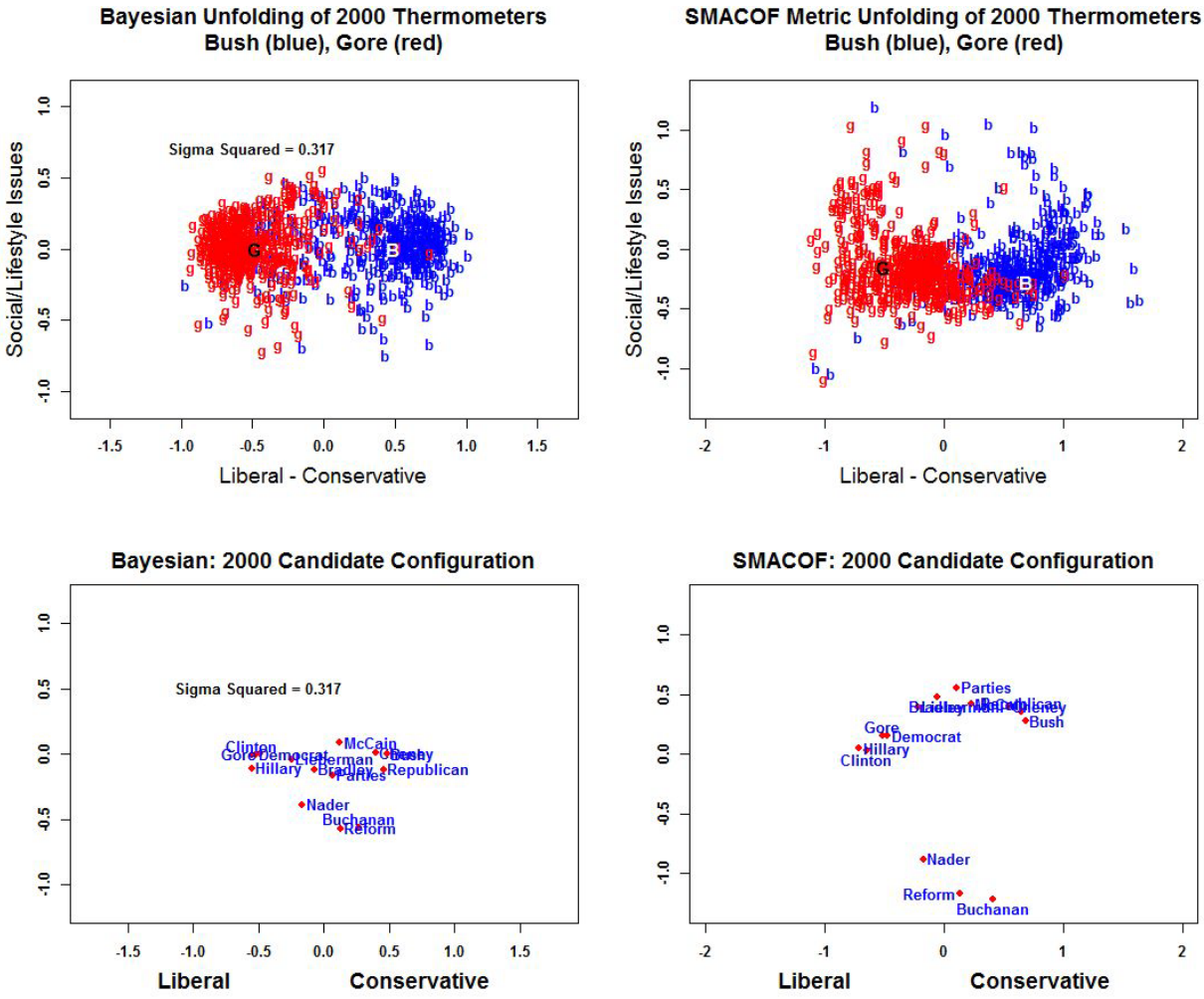

Figure 6: 2004 Thermometer Bayesian vs. SMACOF Unfolding

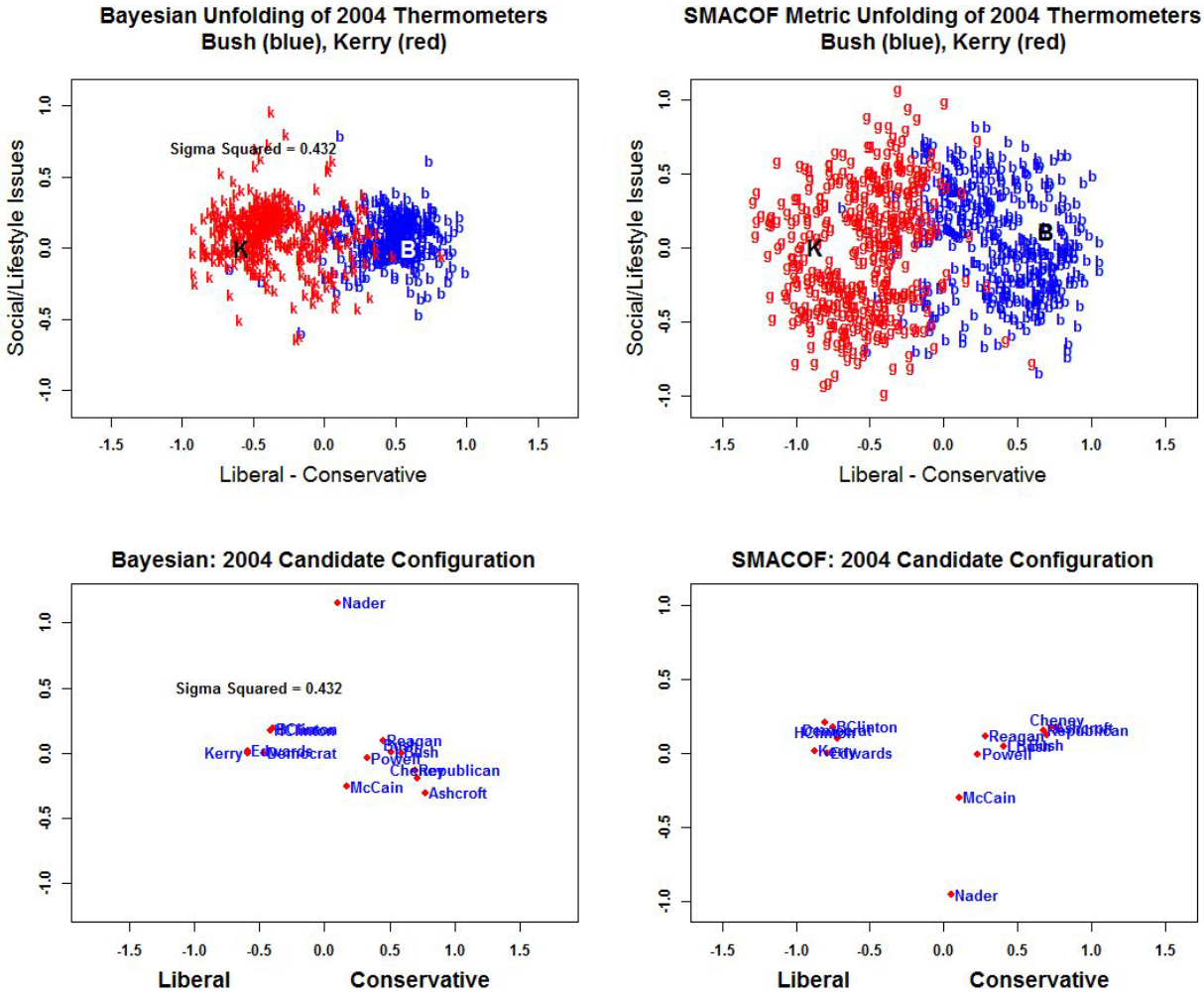

Figure 7: Major Candidates 1968 Presidential Election

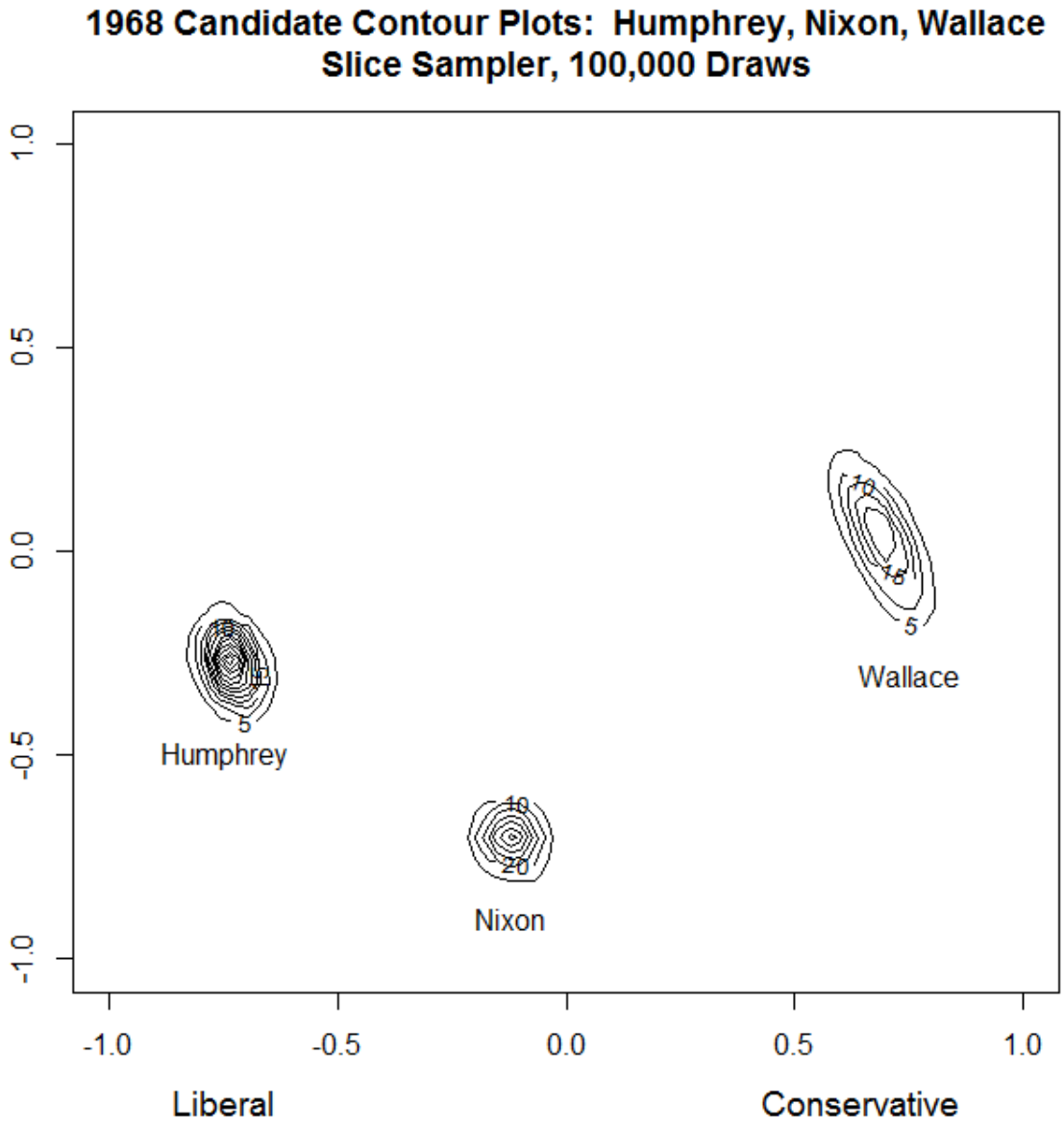

Figure 8: 1968 NES Respondent 2