From *SIAM News*, Volume 38, Number 7, September 2005

Agreement Among Supreme Court Justices: Categorical vs. Continuous Representation

By Lawrence Hubert and Douglas Steinley

On Saturday, July 2, the lead headline in *The New York Times* read as follows: "O'Connor to Retire, Touching Off Battle Over Court." Opening the story attached to the headline, Richard W. Stevenson wrote, "Justice Sandra Day O'Connor, the first wo-man to serve on the United States Supreme Court and a critical swing vote on abortion and a host of other divisive social issues, announced Friday that she is retiring, setting up a tumultuous fight over her successor."

Our purpose here is not to expand on this statement, or even to argue that, with the an-nouncement of the nominee, John Roberts, the word "tumultuous" may no longer be apt. Our interests are in the data set also provided by the *Times* that day, quantifying the (dis)agreement among the Supreme Court justices during the decade they had been together.

In this article, we provide two analyses of these data, and invite others to contribute insights, using whatever methodologies are uppermost in their repertoires for dealing with square and symmetric proximity matrices. And we expect that many students in multivariate statistical analysis classes will be asked to do the same for applied homework projects.

The information in the data set from the *Times* appears in Table 1 in the form of the percentage of non-unanimous cases in which the justices *dis*agreed, from the 1994/95 term through 2003/04. The dissimilarity matrix (in which larger entries reflect less similar justices) is given in the same row and column order as the *Times* data set, with the justices ordered from "liberal" to "conservative":

1: John Paul Stevens (St)
2: Stephen G. Breyer (Br)
3: Ruth Bader Ginsberg (Gi)
4: David Souter (So)
5: Sandra Day O'Connor (Oc)
6: Anthony M. Kennedy (Ke)
7: William H. Rehnquist (Re)
8: Antonin Scalia (Sc)
9: Clarence Thomas (Th)

We present two analyses of the proximity data of Table 1: (a) a unidimensional scaling of the justices, including the estimation of an additive constant that we can apply to the proximities; (b) a hierarchical (or categorical) classification through what is called an "ultrametric," also obtained through a least-squares search strategy. These representations, as a best-fitting unidimensional scale and a best-fitting ultrametric, are generated from methods presented in a forthcoming monograph by Hubert, Arabie, and Meulman, using the available open-source M-files (within a MATLAB environment) that will be provided with this text. The monograph, *The Structural Representation of Proximity Matrices with MATLAB*, is scheduled to appear in 2006 as part of the ASA–SIAM Series on Statistics and Applied Probability.

Unidimensional Scaling

The unidimensional scaling task can be formally phrased as follows: Given the $n \times n$ (in this case, 9×9) proximity matrix $\mathbf{P} = \{p_{ij}\}$ from Table 1, we wish to find an additive con-stant c and a set of coordinates $x_1, \ldots, x_n$ to minimize the least-squares criterion

$$\sum_{i \neq j} \left(p_{ij} + c - |x_j - x_i| \right)^2.$$

The best-fitting result was obtained for the following set of coordinates (with the carets (^) indicating that the values are the best estimates): $\widehat{x}_{\mathrm{St}} = -.346$; $\widehat{x}_{\mathrm{Br}} = -.216$; $\widehat{x}_{\mathrm{Gi}} = -.200$; $\widehat{x}_{\mathrm{So}} = -.177$; $\widehat{x}_{\mathrm{Oc}} = .062$; $\widehat{x}_{\mathrm{Ke}} = .113$; $\widehat{x}_{\mathrm{Re}} = .160$; $\widehat{x}_{\mathrm{Sc}} = .302$; $\widehat{x}_{\mathrm{Th}} = .302$. The additive constant $\hat{c} = -.218$. Notice that the coordinates are ordered exactly as the *Times* ordered the justices in Table 1 (and also that, without loss of generality, the sum of the estimated coordinate values is set to zero). Normalizing the least-squares criterion, we obtain what is usually called a "variance-accounted-for" (VAF) measure:

$$\mathrm{VAF} = 1 - \frac{\sum_{i \neq j} \left(p_{ij} - [\,|\hat{x}_j - \hat{x}_i| - \hat{c}\,] \right)^2}{\sum_{i \neq j} \left(p_{ij} - \bar{p} \right)^2},$$

	St	Br	Gi	So	Oc	Ke	Re	Sc	Th
1 St	.00	.38	.34	.37	.67	.64	.75	.86	.85
2 Br	.38	.00	.28	.29	.45	.53	.57	.75	.76
3 Gi	.34	.28	.00	.22	.53	.51	.57	.72	.74
4 So	.37	.29	.22	.00	.45	.50	.56	.69	.71
5 Oc	.67	.45	.53	.45	.00	.33	.29	.46	.46
6 Ke	.64	.53	.51	.50	.33	.00	.23	.42	.41
7 Re	.75	.57	.57	.56	.29	.23	.00	.34	.32
8 Sc	.86	.75	.72	.69	.46	.42	.34	.00	.21
9 Th	.85	.76	.74	.71	.46	.41	.32	.21	.00

Table 1. *Dissimilarities among the nine current Supreme Court justices.*

where $\bar{p}$ is the mean off-diagonal proximity measure in **P**; the value we observe is 98.0%. In other words, the unidimensional scaling provides a very good representation for the data in Table 1; the O'Connor coordinate, .062, is the coordinate closest to zero and the median of the nine coordinate values over the justices. It would be possible to provide a table for reconstructing the dissimilarities among the justices, using the values $\{|\hat{x}_j - \hat{x}_i| - \hat{c}\}$; direct comparison with the data in Table 1 would reflect the high quality of the reconstruction.

Although the O'Connor coordinate is the median value among the nine locations, a graphical representation (Figure 1) clearly shows that she groups very closely with Kennedy and Rehnquist; the gap between O'Connor and Souter, the closest colleague to her left, is rather large. In choosing the next justice, then, an equivalence would be more toward a Kennedy/Rehnquist conservative than toward a Scalia/Thomas conservative.

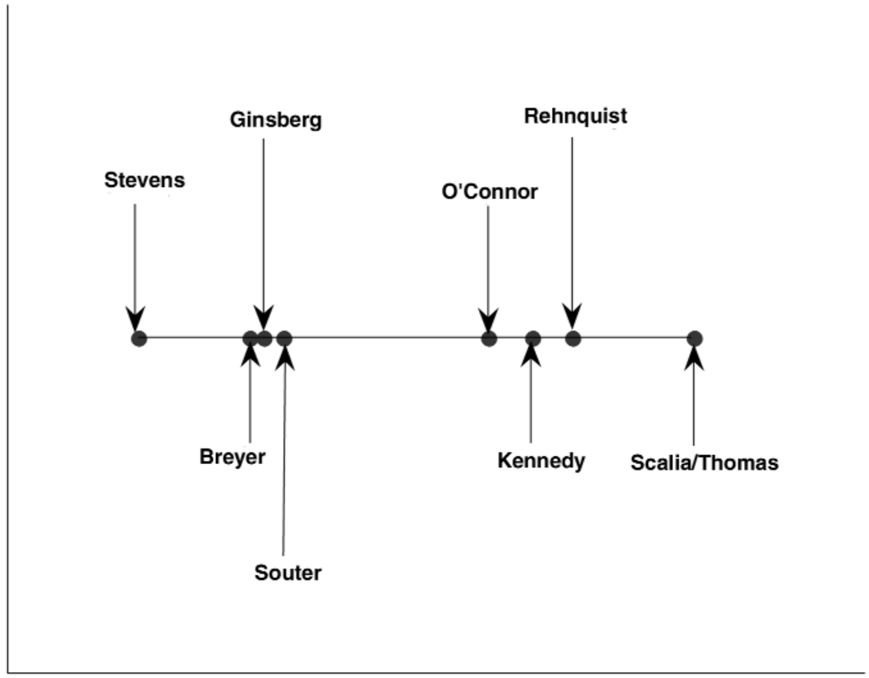

Figure 1. *Unidimensional scaling of the nine Supreme Court justices (based on the coordinates given in the text).*

Hierarchical Classification (Clustering)

Rather than relying on a set of coordinates (and their absolute differences) to represent the elements in a proximity matrix, a best-fitting ultrametric constructs a second matrix to approximate **P** (say, $\mathbf{U} = \{\hat{u}_{ij}\}$), minimizing the least-squares criterion

$$\sum_{i \neq j} \left(p_{ij} - \hat{u}_{ij} \right)^2,$$

where the entries in **U** satisfy the ultrametric inequality $\hat{u}_{ij} \leq \max\{\hat{u}_{ik}, \hat{u}_{kj}\}$ for all i, j, and k. This best-fitting ultrametric can be found via a heuristic search method using iterative projection onto closed convex cones defined by the ultrametric inequality conditions. (This is documented in some detail in the monograph mentioned earlier.)

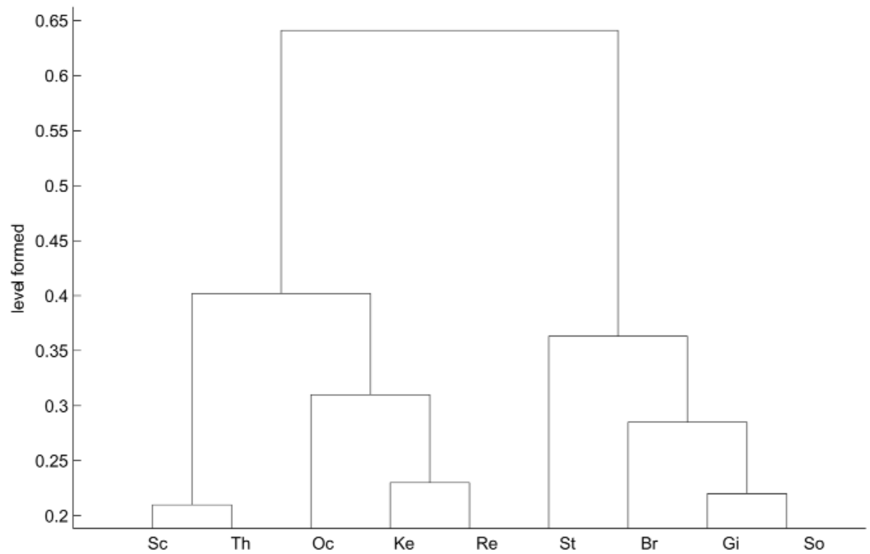

Figure 2. *Dendrogram representation for the best-fitting ultrametric.*

An ultrametric, as thoroughly discussed in the classification literature, induces a partition hierarchy by successive binary subdivision, proceeding from a trivial partition containing all objects within a single class to a second trivial partition having a class for each separate object; the $n - 1$ distinct values that the ultrametric can take on characterize the levels at which the partitions could be considered formed, and also indicate the heights of the nodes in the dendrogram of Figure 2 and the hierarchical organization of the partitions. In the case of the data from Table 1, the best-fitting ultrametric is defined by the eight nonzero distinct values that indicate how the hierarchical sequence of partitions is constructed:

Partition	Level Formed
{Sc, Th, Oc, Ke, Re, St, Br, Gi, So}	.641
{Sc, Th, Oc, Ke, Re}, {St, Br, Gi, So}	.402
{Sc, Th}, {Oc, Ke, Re}, {St, Br, Gi, So}	.363
{Sc, Th}, {Oc, Ke, Re}, {St}, {Br, Gi, So}	.310
{Sc, Th}, {Oc}, {Ke, Re}, {St}, {Br, Gi, So}	.285
{Sc, Th}, {Oc}, {Ke, Re}, {St}, {Br}, {Gi, So}	.230
{Sc, Th}, {Oc}, {Ke}, {Re}, {St}, {Br}, {Gi, So}	.220
{Sc, Th}, {Oc}, {Ke}, {Re}, {St}, {Br}, {Gi}, {So}	.210

The VAF for the ultrametric representation (73.7%) is less adequate than that of the unidimensional scale. A dendrogram for this best-fitting ultrametric is shown in Figure 2. Notice that the order in which the justices are given is different from that in the unidimensional scaling; although we could have maintained the ordering of the unidimensional scaling, we chose not to so as to emphasize the basic “unorderedness” of the clusters implied by the construction of the ultrametric.

Conclusion

It appears that agreement among the Supreme Court justices is better represented as a unidimensional scaling than as a categorical structure defined by a hierarchy of partitions through an associated ultrametric. In terms of Justice O’Connor, despite her placement in the middle of the scaling, the analysis reveals a major tilt toward the conservative end: Her coordinate value is very close to those for Kennedy and Rehnquist, and very discrepant from that for Souter, the colleague to her immediate left.

We did not analyze other dissimilarity matrices that might have resulted from different disaggregations, such as by the type of case under consideration or the plurality of the vote (e.g., in 5-to-4 resolutions). In addition, we have made no particular psychological interpretation of the strong unidimensionality observed in our aggregate analyses, and we have not speculated about the underlying decision mechanisms. As mentioned at the beginning of this article, we invite others to pursue these and other analyses and interpretations.

Lawrence Hubert (lhubert@uiuc.edu) is a professor of psychology and of statistics at the University of Illinois, Urbana–Champaign. Douglas Steinley (steinleyd@missouri.edu) is an assistant professor of psychological sciences at the University of Missouri, Columbia.