Decision Analysis, Social Choice Theory, and Model Selection

Andrew A. Neath
Department of Mathematics and Statistics,
Southern Illinois University Edwardsville

Joseph E. Cavanaugh
Department of Biostatistics,
The University of Iowa

Adam G. Weyhaupt
Department of Mathematics and Statistics,
Southern Illinois University Edwardsville

1 A Decision Problem

Consider the following formulation of a decision problem. Let $\delta_1, \ldots, \delta_L$ denote L distinct courses of action. Let $\Delta(\delta_l, F_*)$ denote the risk from taking action δ_l when the true state of nature is F_*. The preferred course of action under F_* is the one for which the risk is smallest. The problem becomes interesting when the true state of nature F_* is unknown. However, the problem is hopeless without at least some information on the true state F_*, either from data, expert opinion, or some other source. Suppose the information on F_* can be quantified via a probability simulation. Let $F_*^{(1)}, \ldots, F_*^{(N)}$ denote simulated representations of the underlying truth. Let $\Delta_l^{(j)} = \Delta\left(\delta_l, F_*^{(j)}\right)$ denote the risk from taking action δ_l when $F_*^{(j)}$ is the true state. Under each representation of the truth, we are able to determine an order and degree of preference for the available courses of action. A decision is reached by invoking a procedure for turning the varying preferences on the courses of action under the multiple representations of the truth into an ultimate selection. In this article, we look to analyze the process of using the available information in reaching a decision.

2 Social Choice Theory

In the problem under our consideration in this article, we face a decision on the best course of action. We are able to judge each course of action with respect to each of N simulated representations of the true state. The information available to us in the decision problem can be displayed as

$$
\begin{array}{cc|cccc}
 & & & \text{action} & & \\
 & & \delta_1 & \delta_2 & \dots & \delta_L \\
\hline
 & F_*^{(1)} & \Delta_1^{(1)} & \Delta_2^{(1)} & \dots & \Delta_L^{(1)} \\
\text{truth} & \vdots & \vdots & \vdots & \dots & \vdots \\
 & F_*^{(N)} & \Delta_1^{(N)} & \Delta_2^{(N)} & \dots & \Delta_L^{(N)}
\end{array}
$$

The problem of deciding on the best course of action can be analyzed as a voting problem. In a voting problem, each voter has an individual viewpoint on the merits of each candidate. An election is a summary of voter preference. The field of social choice addresses the problem of finding procedures that will turn a collection of individual preferences into an overall ranking. Let $V = \{1, 2, \dots, N\}$ be a finite set of voters, and let $C = \{C_1, \dots, C_L\}$ be a finite set of candidates. Let $B = \{B_1, \dots, B_N\}$ denote the ballots, where B_j is an L-tuple formed as a permutation of the elements in C. For example, the ballot (C_3, C_1, C_2) indicates a voter that prefers candidate C_3 to C_1 to C_2. A social choice procedure is a function which takes the N ballots as input and returns a single output which represents a selection from the candidates in C.

Social choice theory, or voting theory, addresses the goal of finding good procedures for turning individual preferences into a group decision. This is precisely the aim of a decision problem under the framework considered. The available courses of action will represent the candidates in an election. The simulated truths will represent the individual voters. The risk $\Delta_l^{(j)}$ will represent how the j^{th} voter views the l^{th} candidate. Thus, the analysis of a decision making process can be achieved through a study of social choice theory.

The most celebrated work in social choice theory is Arrow's Impossibility Theorem. In a sense, Arrow proved that all social choice procedures are imperfect. Regardless of the fact that no single social choice procedure can be considered as "best", we will use an application to illustrate how the various social procedures, although imperfect, each offer an interesting perspective on the decision problem.

3 Model Selection

Model selection is a special case of a decision problem. Here, we look to determine the theory, or model, which can best explain the underlying truth behind an observed experimental result. Let $\delta_1, \dots, \delta_L$ denote a candidate collection of approximating models. Let F_* denote the true model responsible for generating the experimental data. A discrepancy function $\Delta(\delta_l, F_*)$, analogous to the risk function introduced in Section 1, measures how closely each candidate model approximates the truth. The true model is unknown. But suppose, as before, we can repeatedly simulate representations of the truth, denoted as $F_*^{(1)}, \dots, F_*^{(N)}$. Let $\Delta_l^{(j)} = \Delta\left(\delta_l, F_*^{(j)}\right)$. The discrepancy $\Delta_l^{(j)}$ can be thought of as a measure of the disparity between candidate model δ_l and the j^{th} representation of the truth.

As an example of model selection, consider a linear regression variable selection problem. Let data $y_1, \dots, y_n$ represent independent responses at levels $x_1, \dots, x_n$, respectively, of a p-dimensional vector of predictor variables. The observed levels of the predictor variables can be described in a design matrix X_*. Assume the true distribution corresponds to the linear regression model

$$y = X_* \beta_* + e_*, \quad e \sim N\left(0, \sigma_*^2 I\right).$$

Here, X_* is an $n \times p$ full rank matrix, β_* is a $p \times 1$ coefficient vector, and σ_*^2 is the common error variance. Consider an approximating model δ_l having the same fundamental structure as the true model, but based on a subset of the predictor variables:

$$y = X_l \beta_l + e_l, \quad e_l \sim N\left(0, \sigma_l^2 I\right)$$

where X_l is an $n \times p_l$ design matrix with column space $C(X_l) \subseteq C(X)$.

The Gauss discrepancy is a reasonable judge of an approximating model in the regression setting. The Gauss discrepancy, equivalent to mean squared error, can be decomposed into a variance and bias as

$$\Delta(M_l, M_*) = p_l + \frac{\left\| X_* \beta_* - H_l X_* \beta_* \right\|^2}{\sigma_*^2}$$

where H_l is the projection matrix for column space $C(X_l)$.

We can adopt a Bayesian approach to simulating representations of the truth. The uncertainty inherent to the specification of the true model is characterized through the uncertainty associated with the parameters β_* and σ_*^2. In an effort to stay objective, we can take a noninformative prior on these parameters, although it is not necessary to follow this convention if good prior information is available. After observing data y from the true regression model, the posterior distribution becomes

$$\begin{aligned} \beta_* \mid \sigma_*^2, y &\sim N_{p+1}\left(\widehat{\beta}_*, \sigma^2 \left(X_*' X_*\right)^{-1}\right) \\ \sigma_*^2 \mid y &\sim \frac{RSS_*}{\chi^2(n-p)} \end{aligned} \tag{1}$$

where $\widehat{\beta}_*$ is the least squares estimate of β_* under the full model, and RSS_* is the residual sum of squares under the full model. One can easily simulate a $\left(\beta_*^{(j)}, \sigma_*^{2(j)}\right)$ from the posterior distribution, resulting in a version of the true model $F_*^{(j)}$. For each candidate model, calculate

$$\Delta_l^{(j)} = p_l + \frac{\left\| X_* \beta_*^{(j)} - H_l X_* \beta_*^{(j)} \right\|^2}{\sigma_*^{2(j)}} \tag{2}$$

as a comparison of the l^{th} candidate model to the j^{th} simulated version of the truth.

4 An Application

We will use the Hald data, a well-known application of variable selection in linear regression, as an illustration of a decision problem. Inputs x_1, x_2, x_3, x_4 measure the percentage composition of four ingredients in cement concrete. The design matrix X_* does not make for a perfect mixture experiment, but a degree of collinearity does exist among the predictor variables. Response y measures the heat evolved in calories per gram of concrete. The problem is small (sample size $n = 13$, full model size $p = 5$), yet has interesting features for analyzing the decision making process. The candidate collection includes all regression models formed by taking a subset of the four predictor variables. An initial look at the data reveals that several of these candidate models receive no support from the data. In what follows, our attention is restricted to the seven models in serious contention.

We generated $N = 5000$ representations of the true model using the Bayesian posterior distributions in (1). For each representation and each candidate model, we compute $\Delta_l^{(j)}$ as in (2). We thus have a ranking of preference for each of the candidates from each representation. We will now provide an overview of some social choice procedures, along with a discussion on how each procedure is useful for analyzing the decision problem.

4.1 Plurality

Plurality is the most familiar of the social choice procedures. Plurality selects the candidate with the greatest number of first place votes. The most serious weakness to plurality is the possibility of vote splitting. This phenomenon is more likely to occur when the number of alternatives is large and when some candidates are ideologically similar. For a decision problem, each representation of the truth "votes" for the candidate model judged to be nearest to that version of the truth as measured by the discrepancies $\Delta_l^{(j)}$ in (2). The method of plurality evaluates the candidate models in order of the number of times each model is nearest to a simulated truth. The candidate models, denoted by which input variables are included, can be listed according to plurality as

model	votes	proportion
12	1270	.2504
123	1017	.2034
134	731	.1462
1234	557	.1114
14	517	.1034
124	472	.0944
234	436	.0872

Model 1,2 is at the top of the list. It is of interest to note that Model 1,2,4, is hurt by a vote splitting type effect. Predictors X_2 and X_4 are highly correlated ($r = -.973$). In representations where Model 1,2,4 is evaluated well, models without the redundancy of both X_2 and X_4 are judged similarly, resulting in relatively few scenarios for which Model 1,2,4 is evaluated as best in the candidate class. This idea can be further explored if models are compared pairwise.

4.2 Condorcet / Copeland

The Condorcet method outputs the candidate which pairwise beats or ties every other alternative in a head to head vote. If no candidate beats every other candidate head to head, then the Condorcet method does not output a social choice. Copeland's method counts the number of pairwise wins for each candidate and ranks the candidates in order of the most wins. If a Condorcet winner emerges then clearly that candidate wins under Copeland as well. But a property of Copeland's method is the output of a winner even in situations where a Condorcet winner does not emerge.

Pairwise comparisons for the top candidate models can be displayed as

	124	**123**	**134**	**1234**	**12**
1234	.6682	.6666	.5658	-	.3766
12	.6160	.6136	.5782	.6234	-
14	.8244	.8222	.8022	.8094	.6522
123	.5108	-	.4040	.3334	.3864
124	-	.4892	.3466	.3318	.3840
134	.6534	.5960	-	.4342	.4218
234	.8512	.8156	.8968	.8220	.7196
wins	6	5	4	3	2

The proportion of wins for one model over the others in the candidate class are found by reading down a column. We can see the effect of the vote splitting alluded to earlier. Model 1,2,4, evaluated low by a plurality vote, is the winner according to Copeland's method. Furthermore, Model 1,2,4 is a Condorcet winner in that Model 1,2,4 is preferred head to head over every other model in the candidate class.

4.3 Borda count / rank method

The Borda count is a positional system; candidates are awarded points based on their ranking on each ballot. A Borda count evaluates candidates by the total number of points across all ballots. Evaluating candidates by their average rank across all ballots is equivalent to a Borda count.

Rankings for the top candidate models can be displayed as

	124	**123**	**134**	**1234**	**12**
1st	472	1017	731	557	1270
2nd	1516	1301	821	654	345
3rd	1659	995	777	830	379
4th	949	708	1092	1272	567
5th	321	378	1514	1369	608
6th	83	549	65	285	925
7th	0	52	0	33	906
avg	2.88	3.00	3.41	3.65	4.06

Each column contains the number of ballots giving that rank to the particular candidate model, as well as the average rank over all ballots for that candidate model. The final ordering of preference for the models is the same as the ordering from Copeland's method. (A Condorcet winner is necessarily a Borda count winner, although the converse does not hold.) Model 1,2,4 is the selected model based on average ranking.

Based on first place votes alone as in plurality, Model 1,2 is selected. By comparison, Model 1,2,4 receives relatively few first place votes. One can argue in favor of Model 1,2 over Model 1,2,4 in that the model with the most first place votes is the model deemed most likely to be the one which minimizes the underlying discrepancy. Based on average rank, Model 1,2,4 is selected. One could then argue in favor of Model 1,2,4 over Model 1,2. Although relatively few scenarios have Model 1,2,4 ranked first, a substantial proportion have this model ranked near the top. By comparison, there are many scenarios in which Model 1,2 is ranked near the bottom of the listed candidate models.

4.4 Hare system / single transferable vote

In a Hare system, or single transferable vote (STV) system, the candidate with the fewest number of first place votes is eliminated in the first round. Ballots for the eliminated candidate have their votes transfered to their respective second choices. The procedure continues eliminating candidates with the fewest votes, and transferring those votes to the individual voters next choice, until only one candidate remains. Results for the top candidate models can be displayed as

	123	**134**	**12**	**1234**	**14**	**124**
1	1017	731	1270	557	517	472
2	1017	933	1276	757	543	474
3	1271	1123	1293	765	548	0
4	1473	1480	1302	781	0	0
5	1809	1726	1465	0	0	0
6	2980	2020	0	0	0	0
rnd	-	6	5	4	3	2

Each column gives the number of votes for that candidate in each round of voting, and the round of elimination for each candidate.

Now, Model 1,2,3 is selected. A STV system attaches more importance to first place votes than does the average rank method. Model 1,2,4 does not have enough first place votes to last very long in the competition. A STV system differs from plurality in that a model requires a secondary level of support. For example, Model 1,2 is eliminated because of a lack of transfered votes to support its first place votes.

5 Concluding Remarks

Decision analysis seeks to understand the approach one takes in reaching a decision based on limited information. In this article, we have shown how a decision problem can be cast in the framework of a voting problem. Thus, social choice theory becomes applicable to decision analysis.

Arrow's Impossibility Theorem states that there is no "best" social choice procedure. A decision problem then faces the same ambiguity. The procedure one uses for reaching a decision depends on the needs of the decision maker. In the variable selection application, an investigator has a reasonable justification for deciding on any of the models that came out on top of a social choice procedure. Model 1,2 represents a high risk / high reward decision in that there is a good chance this decision will turn out best, but also a good chance that this decision will fare poorly. Model

1,2,4 represents a safe decision in that there is a small chance that this decision will fare poorly, although there is also only a small chance this decision will turn out best. Model 1,2,3 represents a compromise.

Arrow's Impossibility Theorem has not slowed the field of social choice. On the contrary, the study and debate of social choice intensified after it became clear that no single best solution exists. Decision scientists can benefit from the interest in social choice procedures. The use of social choice theory adds an interesting perspective to the analysis of a decision problem.

6 Further Reading

Arrow, K. (2002). *Collected Papers of Kenneth J. Arrow*. Cambridge: Belknap Press.

Saari, D.G. (2001). *Decisions and Elections: Explaining the Unexpected*. Cambridge: University Press.

Thompson, G. (2010). Keeping things in proportion: How can voting systems be fairer? *Significance*, 7, 128-132.