Voting, Ranking, and Preference Models

Carolina Sawyer & Michael Pearce (Mentor)

During my project for the UW Statistics and Probability Association's Directed Reading Program, I learned about voting, ranking, and preference data. Each week my mentor Michael and I discussed these topics beginning with theory and eventually focusing on applying and coding the models I learned. For my final project, I created an R Shiny app which can be used to model and visualize a user's ranking data.

In the first few weeks, Michael and I discussed what preference data is and how it is used. Preference data occurs when an individual compares items and expresses a preference for one item over another. These can be pair-wise comparisons, or a set of items being compared and ranked. Preference data is used in voting systems, search algorithms, and recommender systems, for example. Many services like Netflix use preference data to offer personalized product recommendations to users. One type of preference data we discussed in depth is voting and different voting systems. Single-member district plurality voting is most used in the U.S., in which the winner is the candidate with the most votes. In contrast, the single-transferable vote has voters rank candidates. In case a voter's first choice is eliminated or has an excess of votes needed to win, their vote is transferred to their second choice. Instant runoff is another ranked voting system, where a similar process is repeated until a candidate has the majority of votes. I never realized there are so many types of voting systems before, so I thought it was particularly interesting learning about the benefits and drawbacks of each. It helped me think about the importance of how preference data is collected and assessed.

Next, we moved onto modeling preference data. We first discussed different methods for measuring the distance between rankings. For example, Spearman's footrule measures the sum of differences between pairwise object ranks while the Hamming distance sums how often two objects are assigned different places. I realized there are many ways to interpret a set of rankings and depending on the type of data, certain metrics may be more appropriate.

Following that, I continued to learn about several preference models and coding them in R. Two models we discussed in detail were the Plackett-Luce and Mallows models. The Plackett-Luce model provides coefficients describing how likely an object is to be picked first. The object with the largest coefficient is most likely to be picked and a consensus ranking can be formed by ordering the objects based on their coefficients. Depending on how close the coefficients are, we can get an idea of how strong the consensus is. The Mallows model similarly gives a consensus ranking as well as a theta value describing the strength of the consensus. A small theta value means there is less consensus in the group, while a large theta value means there was strong consensus in the group.

This led to my final project, where I created a Shiny app that uses the Plackett-Luce and Mallows models to describe rankings. Within the app, a user can upload their own data and view several tabs with information summarizing the data. There is an overall summary tab with a graph displaying each ranking place and how often an object was chosen in that place. Another tab contains information from the Mallows model and a graph with each ranking's distance from consensus. The last tab displays the Plackett-Luce model's coefficients and consensus ranking.

Overall, I have gained so much new knowledge about voting, ranking, and preference data including experience with coding and modeling ranking data. I’m grateful to my mentor, Michael, for explaining things that were harder for me to understand and helping me along the way. I’ve really enjoyed what I’ve learned so far and am excited to keep working on my app and continue learning more about modeling preference data in the future.