1

Mixed Membership Models for Rank Data: Investigating Structure In Irish Voting Data

Isobel Claire Gormley

School of Mathematical Sciences & Complex and Adaptive Systems Laboratory, University College Dublin, Dublin 4, Ireland

Thomas Brendan Murphy

School of Mathematical Sciences & Complex and Adaptive Systems Laboratory, University College Dublin, Dublin 4, Ireland

CONTENTS

A mixed membership model is an individual level mixture model where individuals have partial membership of the profiles (or groups) that characterize a population. A mixed membership model for rank data is outlined and illustrated through the analysis of voting in the 2002 Irish general election. This particular election uses a voting system called proportional representation using a single transferable vote (PR-STV) where voters rank some or all of the candidates in order of preference. The data set considered consists of all votes in a constituency from the 2002 Irish general election. Interest lies in highlighting distinct voting profiles within the electorate and studying how voters affiliate themselves to these voting profiles. The mixed membership model for rank data is fitted to the voting data and is shown to give a concise and highly interpretable explanation of voting patterns in this election.

1.1 Introduction

Mixture models are a well established tool for statistical model-based clustering of data [39, 16]. Mixture models describe a population as a finite collection of homogeneous groups, each of which is characterized by a specific probability density. While based on a similar concept, mixed membership (or grade of membership (GoM)) models allow every individual have partial membership of each of the groups that characterize the population. Thus mixed membership models provide a method for model-based soft clustering of data. The mixed membership (or GoM) model for multivariate categorical data is developed in [13] and [3] and this model has been used in a number of applications including [12, 11, 2] amongst others.

Rank data arise when a set of judges rank some (or all) of a set of objects. Rank data emerge in many areas of society; the final ordering of athletes in a race, league tables, the ranking of relevant results by internet search engines and consumer preference data provide examples of such data. In this chapter, a mixed membership model for rank data that was originally developed in [24] is described and applied to the problem of finding structure in Irish voting data.

The Irish electoral system uses a voting system called proportional representation using a single transferable vote (PR-STV). In this system, voters rank some or all of the candidates in order of preference. When drawing inferences from such data, the information contained in the different preference levels must be exploited by the use of appropriate modeling tools. An illustration of the mixed membership model for rank data methodology is provided through an examination of voting data from the 2002 Irish general election. Interest lies in highlighting voting profiles that occur within the electorate. The mixed membership model provides the scope to examine if and how voters exhibit mixed membership by sharing preference behavior described by more than one of these voting profiles.

A latent class representation of the mixed membership model for rank data is used for model fitting within the Bayesian paradigm. A Metropolis-within-Gibbs sampler is necessary to provide samples from the posterior distribution. Model selection is achieved using the Deviance Information Criterion (DIC) and the adequacy of model fit is assessed using posterior predictive checks.

The chapter proceeds as follows: in Section 1.2 the Irish voting system and details surrounding the 2002 Irish general election are outlined. The Plackett-Luce model for rank data is employed in this application as the rank data model; this model and other rank data models are discussed in Section 1.3.1. The specification of the mixed membership model for rank data follows in Section 1.3.2. Estimation of the mixed membership model for rank data is outlined in Section 1.4.1. The question of model choice is addressed in Section 1.4.2. The application of the mixed membership model for rank data to 2002 Irish general election data is given in Section 1.5. The article concludes in Section 1.6 with a discussion of the methodology.

1.2 The 2002 Irish general election

Dáil Éireann is the main parliament in the Republic of Ireland and it has one hundred and sixty six members. Members (called Teachtaí Dála or TDs) are elected to the Dáil through a general election which must take place at least every five years. On the 17th of

May 2002, a general election was held to elect the 29th Dáil and candidates ran in forty two constituencies. Each constituency elected either three, four or five candidates where the number of candidates to be elected is determined by the population of the constituency; the speaker of the house (Ceann Comhairle) from the previous parliament was automatically returned as elected in their constituency. The outgoing government consisted of a Fianna Fáil and Progressive Democrat coalition with Fianna Fáil having seventy seven seats and the Progressive Democrats having four seats. Thus, the outgoing government was a minority government who relied on a number of independent TDs for support. After the election, a coalition government involving Fianna Fáil and the Progressive Democrats was formed again but returning with a majority and holding eighty one and eight seats, respectively. This was the first time that a government had been returned in an Irish general election in thirty years. Extensive descriptions of the 2002 election are provided by [31, 50, 18, 38].

In the 2002 general election, a trial was conducted in three constituencies (Dublin North, Dublin West and Meath) where electronic voting was introduced. The voting data from these three constituencies was made publicly available and these data provide an unprecedented insight into the voting in Irish elections, beyond what had previously been available in poll data. The data from the Dublin North constituency is analyzed because it contained a particularly diverse range of candidates and thus the data would be expected to contain interesting voting behavior.

In 2002, the Dublin North constituency consisted of an electorate of 72353 and four TDs were to be elected from this constituency. A total of 43942 people voted and twelve candidates ran for election. Fianna Fáil, who were the largest political party at the time, ran three candidates, Fine Gael who were the largest opposition party ran two candidates, the Labour, Green, Sinn Féin parties ran one candidate each and more minor parties like the Socialist, Christian Solidarity and Independent Health Alliance parties also ran one candidate each. One independent candidate ran for election and the Progressive Democrats did not run any candidate in Dublin North. Four of the candidates were incumbent candidates from the 28th Dáil but where Seán Ryan (Labour) was elected to the 28th Dáil through a by-election after the resignation of Ray Burke (Fianna Fáil) from his seat during the 28th Dáil.

Irish general elections employ an electoral system known as the proportional representation by means of a single transferable vote (PR-STV) system. Under this electoral system a voter ranks, in order of his/her preference, some or all of the electoral candidates on a ballot form. The votes are totalled through a series of counts, where candidates are eliminated, their votes are distributed, and surplus votes are transferred between candidates. A detailed introduction to the PR-STV voting system in an Irish context is given in [46] and a good overall comparison of different voting systems is given by [15, 19].

Details of the counting and transfer of votes in the Dublin North constituency are shown in Table 1.1. The total valid poll was 43942, so the number of votes required to guarantee election (called the droop quota) was 8789 votes. In the first count the number of first preferences for each candidate are counted. If no candidate exceeds the droop quota, then the lowest candidates are eliminated and their votes are distributed using their next available preferences. If a candidate is elected by exceeding the droop quota, then their surplus votes (the amount by which they exceed the droop quota) are distributed using their next available preferences. The procedure of eliminating low candidates and distributing surpluses continues until either four candidates have exceeded the droop quota or only four candidates remain. In Dublin North, two candidates reached the quota and two were elected without reaching the quota. The four candidates elected were also the four candidates with the highest number of first preferences, but this does not necessarily happen.

TABLE 1.1
The counting and transfer of votes in the Dublin North constituency in the 2002 Irish general election. The incumbent candidates are marked with an asterisk. The point at which each candidate was elected is marked in bold.

Candidate	Party	Count							
(Abbreviation)		1	2	3	4	5	6	7	8
Trevor Sargent*	Green	7294	7380	7678	7818	8118	**9785**	8789	8789
(Sa)			*+86*	*+298*	*+140*	*+300*	*+1667*	*-996*	
Seán Ryan*	Labour	6359	6407	6535	6665	6847	8578	**9128**	9128
(Ry)			*+48*	*+128*	*+130*	*+182*	*+1731*	*+550*	
Jim Glennon	Fianna Fáil	5892	5945	6028	6152	6294	6511	6598	**8640**
(Gl)			*+53*	*+83*	*+124*	*+142*	*+217*	*+85*	*+2044*
G V Wright*	Fianna Fáil	5658	5707	5739	5777	5868	6139	6249	**8617**
(Wr)			*+49*	*+32*	*+38*	*+91*	*+271*	*+110*	*+2368*
Clare Daly	Socialist	5501	5551	5730	5796	6244	6590	6772	7523
(Dy)			*+53*	*+179*	*+66*	*+448*	*+346*	*+182*	*+751*
Michael Kennedy	Fianna Fáil	5253	5309	5368	5422	5532	5732	5801	
(Ke)			*+56*	*+59*	*+54*	*+110*	*+200*	*+69*	*-5801*
Nora Owen*	Fine Gael	4012	4030	4132	4720	4763			
(Ow)			*+18*	*+102*	*+588*	*+43*	*-4763*		
Mick Davis	Sinn Féin	1350	1382	1424	1440				
(Dv)			*+32*	*+42*	*+16*	*-1440*			
Cathal Boland	Fine Gael	1177	1189	1216					
(Bo)			*+12*	*+27*	*-1216*				
Ciarán Goulding	Independents	914	1009						
(Go)	Health Alliance		*+95*	*-1009*					
Eamon Quinn	Independent	285							
(Qu)			*-285*						
David Walshe	Christian	247							
(Wa)	Solidarity Party		*-247*						
Non Transferable			33	92	152	276	607	607	1245
			+33	*+59*	*+60*	*+124*	*+331*		*+638*
Total		43942							

1.3 Model specification

The Dublin North general election voting data possess some unique properties which require careful statistical modeling. A mixed membership model can easily accommodate the differing preferences that voters may have for the candidates. Although a finite mixture model may be used for the same purpose [eg. 22] the finite mixture model needs a large number of mixture components to account for the voting behavior exhibited in the electorate; conversely the mixed membership model can account for different behavior using a relatively small number of profiles. In order to account for the ranked nature of the preference voting data, the Plackett-Luce model for rank data is used.

1.3.1 The Plackett-Luce model for rank data

Under the PR-STV electoral system a voter ranks some or all of the candidates in order of preference. In order to appropriately model such data, a model for rank data is required. A large number of models for rank data have already been developed [4, 36, 43] and these are reviewed in [37]. In this study the Plackett-Luce model [43] is utilized to model the rank nature of the data.

The Plackett-Luce model is parameterized by a 'support' parameter

$$\underline{p} = (p_1, p_2, \ldots, p_N)$$

where N denotes the total number of electoral candidates. Note that $0 \leq p_j \leq 1$ and $\sum_{j=1}^{N} p_j = 1$. The parameter p_j has the interpretation of being the probability of candidate j being ranked first by a voter. The probability of candidate j being given a lower than first preference is proportional to their support parameter p_j. Hence, at preference levels lower than the first the probabilities are re-normalized to provide valid probability values.

Let voter i record the vote $\underline{x}_i = \{c(i,1), c(i,2), \ldots, c(i,n_i)\}$, where n_i is the number of preferences expressed by voter i. The Plackett-Luce model states that the probability of vote $\underline{x}_i$ is given as

$$\begin{aligned} \mathbf{P}\{\underline{x}_i|\underline{p}\} &= \prod_{t=1}^{n_i} \frac{p_{c(i,t)}}{p_{c(i,t)} + p_{c(i,t+1)} + \cdots + p_{c(i,N)}} \\ &= \prod_{t=1}^{n_i} \frac{p_{c(i,t)}}{\sum_{s=t}^{N} p_{c(i,s)}} = \prod_{t=1}^{n_i} q_{it}, \end{aligned} \tag{1.1}$$

where $c(i, n_i+1), \ldots, c(i,N)$ is any permutation of the unranked candidates. Note that the probability of the ranking is conditional on n_i, the number of preferences expressed and it can easily be shown that (1.1) sums to 1 over all $n_i!$ possible permutations of the candidates ranked in the vote $\underline{x}_i$.

1.3.2 The mixed membership model for rank data

Mixed membership models allow every individual in a population have partial membership of each of the profiles that characterize the population; thus, a soft clustering of the population members is achievable. Herein we describe a mixed membership model for rank data as developed by [24].

Under the mixed membership model each voter $i = 1, \ldots, M$ has an associated *mixed membership parameter* $\underline{\pi}_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{iK})$ which is a direct parameter of the model.

The mixed membership parameter $\underline{\pi}_i$ describes the degree of membership of individual i in each of the K profiles which characterize the electorate. Note that $0 \leq \pi_{ik} \leq 1$ and $\sum_{k=1}^{K} \pi_{ik} = 1$ for $i = 1, \ldots, M$. Thus, if individual i is fully characterized by profile k, then $\pi_{ik} = 1$ and $\pi_{ij} = 0$ for $j \neq k$. Additionally, if individual i is characterized by profiles $\mathcal{K} \subset \{1, 2, \ldots, K\}$, then $\pi_{ij} > 0$ for $j \in \mathcal{K}$ and $\pi_{ij} = 0$ for $j \notin \mathcal{K}$.

The mixed membership model for ranked data is formulated as follows. We assume that the probability of voter i ranking candidate j in position t on their ballot is a convex combination of the probability of the voter choosing candidate j in position t as described by each profile, where the weights in the convex combination are equal to the voters mixed membership parameter. That is, the probability of voter i choosing candidate j at preference level t, conditional on voter i's mixed membership parameter $\underline{\pi}_i$ and the profile specific support parameters $\mathbf{p} = (\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_K)$, is given as

$$\mathbf{P}\{c(i,t) = j | \underline{\pi}_i, \mathbf{p}\} = \sum_{k=1}^{K} \pi_{ik} \left[\frac{p_{kj}}{\sum_{s=t}^{N} p_{kc(i,s)}} \right]. \tag{1.2}$$

Additionally, local independence is then assumed between each preference level t, given the mixed membership parameters. Thus, the conditional probability of ranking $\underline{x}_i$ given membership parameter $\underline{\pi}_i$ and support parameters $\mathbf{p}$ is

$$\mathbf{P}\{\underline{x}_i | \underline{\pi}_i, \mathbf{p}\} = \prod_{t=1}^{n_i} \left\{ \sum_{k=1}^{K} \pi_{ik} \left[\frac{p_{kc(i,t)}}{\sum_{s=t}^{N} p_{kc(i,s)}} \right] \right\}$$

and the likelihood function based on the data $\mathbf{x} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_M)$ is therefore

$$\mathbf{P}\{\mathbf{x} | \boldsymbol{\pi}, \mathbf{p}\} = \prod_{i=1}^{M} \prod_{t=1}^{n_i} \left\{ \sum_{k=1}^{K} \pi_{ik} \left[\frac{p_{kc(i,t)}}{\sum_{s=t}^{N} p_{kc(i,s)}} \right] \right\}.$$

Note that under the mixed membership model each voter has partial membership of each profile and mixing takes place at each preference level t rather than at the vote level as would be typical of a rank data mixture model [49, 41, 21, 5, 22, 23]. Modeling rank data in this manner provides a deeper insight to the structure within the electorate by allowing mixing to occur at a finer level. This is a desirable characteristic as it may be restrictive to assume a voter expresses all preferences in their vote as dictated by a single profile; it is likely that a voter may express some preferences in line with the support parameters of one profile, and other preferences in line with the support parameters of other profiles. This is clearer when we look at the latent class representation of the mixed membership model (Section 1.3.2.1).

1.3.2.1 A latent class representation of the mixed membership model

The mixed membership model for rank data can be expressed using a latent class representation in a manner similar to [10]; this representation facilitates efficient inference for the model and it assists with model interpretation. The latent class representation of the mixed membership model for rank data involves augmenting the data for each voter i with categorical latent variables which record the profile that is used by voter i when recording preference level t. The discrete distribution for the latent classes has a functional form that depends on mixed membership parameters $\underline{\pi}_i$ for voter i.

For each voter i, we impute binary latent vectors $\underline{z}_{it} = (z_{it1}, \ldots, z_{itK})$ for $t = 1, \ldots, n_i$ where $\underline{z}_{it} \sim \text{Multinomial}(1, \underline{\pi}_i)$. The value of $\underline{z}_{it}$ records the voting profile that is used by voter i when recording preference level t.

It follows that under the mixed membership model the 'augmented' data likelihood function based on the data $\mathbf{x}$ and the binary latent variables $\mathbf{z}$ is therefore of the form

$$\mathbf{P}\{\mathbf{x}, \mathbf{z}|\boldsymbol{\pi}, \mathbf{p}\} = \prod_{i=1}^{M}\prod_{k=1}^{K}\prod_{t=1}^{n_i}\left\{\pi_{ik}\left[\frac{p_{kc(i,t)}}{\sum_{s=t}^{N} p_{kc(i,s)}}\right]\right\}^{z_{itk}}. \quad (1.3)$$

Employing the latent class representation of the mixed membership model not only allows estimation of the characteristic parameters of each profile but also direct estimation of the mixed membership parameter for each voter, thus achieving a soft clustering of the voters. In addition, the mixed membership of each individual can be further probed to establish which profile is best appropriate for modeling voter i when they are making choice level t.

1.3.3 Prior and posterior distributions

A Bayesian approach is taken when estimating the mixed membership model for rank data and thus the specification of prior distributions for the parameters of the model is required. It is assumed that the mixed membership parameters follow a Dirichlet($\underline{\alpha}$) distribution and that the support parameters follow a Dirichlet($\underline{\beta}$) distribution i.e.

$$\begin{aligned} \underline{\pi}_i &\sim \text{Dirichlet}\left\{\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)\right\} \\ \underline{p}_k &\sim \text{Dirichlet}\left\{\underline{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)\right\}. \end{aligned}$$

The conjugacy of the Dirichlet distribution with the multinomial distribution means the use of a Dirichlet prior is naturally attractive. The use of a Dirichlet prior does however induce a negative correlation structure between parameters. The sensitivity of inferences drawn under the mixed membership model for rank data to this prior specification is considered in [24]. For even moderate sized data sets they found that the posterior inferences were not heavily influenced by the prior specification. In practice the prior parameters are fixed as $\underline{\alpha} = (0.5, \ldots, 0.5)$ and $\underline{\beta} = (0.5, \ldots, 0.5)$ which is the Jeffreys prior for the multinomial distribution [42].

Given these prior distributions and the augmented data likelihood function (1.3) from the mixed membership model for rank data, the posterior distribution based on the data is:

$$\mathbf{P}\{\boldsymbol{\pi}, \mathbf{p}, \mathbf{z}|\mathbf{x}\} \propto \left[\prod_{i=1}^{M}\prod_{k=1}^{K}\prod_{t=1}^{n_i}\left\{\pi_{ik}\left[\frac{p_{kc(i,t)}}{\sum_{s=t}^{N} p_{kc(i,s)}}\right]\right\}^{z_{itk}}\right]\left[\prod_{i=1}^{M}\prod_{k=1}^{K}\pi_{ik}^{\alpha_k-1}\right]\left[\prod_{k=1}^{K}\prod_{j=1}^{N}p_{kj}^{\beta_j-1}\right].$$

This posterior distribution differs from the posterior distribution in the case of the original mixed membership model [13, 14] in the form of the likelihood function. In the original mixed membership model, discrete response variables are treated as independent given the mixed membership parameters. The likelihood function is therefore the product of independent Bernoulli distributions. In the mixed membership model for rank data however, the dependence of choices within a rank response leads to a more complex likelihood function that is the product of terms that share parameter values.

1.4 Model Inference

1.4.1 Parameter Estimation

The mixed membership model for rank data can be efficiently fitted in a Bayesian framework. Due to the structure of the posterior distribution, Markov chain Monte Carlo methods are necessary to produce posterior samples of the model parameters. In particular, a Gibbs sampling step can be used in the algorithm if the full conditional distribution for a model parameter has a tractable form. For most of the model parameters in the mixed membership model for rank data this is indeed the case, however in the case of the support parameters $\mathbf{p}$ it is not the case.

The full conditional distributions of the latent variables $\underline{z}_{it}$ and the mixed membership parameters $\underline{\pi}_i$ are readily available. In particular,

$$\underline{z}_{it} \sim \text{Multinomial}\left\{1, \left(\frac{\pi_{i1}q_{1it}}{\sum_{k'=1}^{K}\pi_{ik'}q_{k'it}}, \frac{\pi_{i2}q_{2it}}{\sum_{k'=1}^{K}\pi_{ik'}q_{k'it}}, \ldots, \frac{\pi_{iK}q_{Kit}}{\sum_{k'=1}^{K}\pi_{ik'}q_{k'it}}\right)\right\},$$

where q_{kit} is defined as in (1.1) for $k = 1, 2, \ldots, K$, $i = 1, \ldots, M$, $t = 1, \ldots, n_i$ and

$$\underline{\pi}_i \sim \text{Dirichlet}\left(\alpha_1 + \sum_{t=1}^{n_i} z_{it1}, \ldots, \alpha_K + \sum_{t=1}^{n_i} z_{itK}\right) \qquad \text{for } i = 1, \ldots, M.$$

In the case of the support parameters, the full conditional distributions are

$$\mathbf{P}\{\underline{p}_k|\boldsymbol{\pi}, \mathbf{x}, \mathbf{z}\} \quad \propto \quad \left[\prod_{i=1}^{M}\prod_{t=1}^{n_i}\left\{\frac{\pi_{ik}p_{kc(i,t)}}{\sum_{s=t}^{N} p_{kc(i,s)}}\right\}^{z_{itk}}\right]\left[\prod_{j=1}^{N} p_{kj}^{\beta_j - 1}\right]. \tag{1.4}$$

Due to the form of the likelihood function based on the rank data, the complete conditional distribution of the support parameters is not readily available for sampling and a Gibbs sampling step cannot be implemented. However, a Metropolis step can be used to sample the support parameters. Thus, a Metropolis-within-Gibbs sampler [7] can be used to sample from the posterior for all model parameters.

In any Metropolis-based algorithm, the rate of convergence of the chain depends on the relationship between the proposal and target distributions. The use of a proposal distribution which closely mimics the shape and orientation of the target distribution provides an improved rate of convergence and good mixing.

The approach taken to construct a proposal distribution starts by examining the logarithm of the full conditional of the support parameter $\underline{p}_k$ (1.4) which is of the form

$$\log \mathbf{P}\{\underline{p}_k|\boldsymbol{\pi}, \mathbf{x}, \mathbf{z}\} \propto \sum_{i=1}^{M}\sum_{t=1}^{n_i} z_{itk}\left\{\log p_{kc(i,t)} - \log \sum_{s=t}^{N} p_{kc(i,s)}\right\} + \sum_{j=1}^{N}(\beta_j - 1)\log p_{kj}.$$

The function $-\log(\cdot)$ is a convex function and thus the term $-\log \sum_{s=t}^{N} p_{kc(i,s)}$ can be approximated (in fact lower bounded) by a hyperplane that is tangent to the function at the currently sampled value of $\underline{p}_k$. The resulting function is the log of a gamma density and this can, in turn, be replaced by the log of a Gaussian density because the shape parameter is typically quite large. Thus, the proposal distribution for p_{kj} emerges as a Gaussian density with mean and variance dependent on the previously sampled values of the model parameters. As the Gaussian distribution extends beyond the $[0, 1]$ interval in which the support parameters lie, proposed values from this surrogate proposal must be suitably normalized.

When estimating parameters via MCMC algorithms some special features of the mixed membership model for ranking data require attention. A fundamental issue in the fitting of any mixture based model within a Bayesian framework is that of label switching. This arises because of the invariance of posterior distribution to permutations in the labeling of the profiles. The methods proposed for dealing with label switching, including [48, 9, 29] need to be considered to avoid this issue. The online relabeling algorithm of [48] was found to be an effective method for handling this issue.

Full details of the Metropolis-within-Gibbs algorithm for fitting this model are given in [24].

1.4.2 Model selection

Another feature of the mixed membership model is the need to infer the model dimensionality, that is, the number of voting profiles (K) needed to appropriately model the electorate. Within the Bayesian paradigm the natural approach would appear to be to base inference on the posterior distribution of K given the data $\mathbf{x}$, $\mathbf{P}\{K|\mathbf{x}\}$. However this posterior can be highly dependent on the model definition and is typically computationally challenging to construct. A comprehensive overview and comparison of model selection criteria within the context of mixed membership models is provided in [30].

In this application of the mixed membership model for rank data, the Deviance Information Criterion (DIC) introduced by [47] is used to choose an appropriate model. The DIC criterion penalizes the posterior mean deviance of a model by the "effective number of parameters". The effective number of parameters is derived to be the difference between the posterior mean of the deviance and the deviance at the posterior means of the parameters of interest. Explicitly for data $\mathbf{x}$ and parameters θ the DIC is

$$DIC = \overline{D(\theta)} + p_D$$

where $D(\theta) = -2\log[\mathbf{P}(\mathbf{x}|\theta)] + 2\log[h(\mathbf{x})]$ is the Bayesian deviance and $h(\mathbf{x})$ is a function of the data only. The effective number of parameters is defined as $p_D = \overline{D(\theta)} - D(\bar{\theta})$. The criterion has an approximate decision theoretic justification. In any case, models with small DIC are preferable to models with large DIC values.

1.5 Application to the 2002 Irish general election

The mixed membership model for rank data was applied to the voting data from the Dublin North constituency in the 2002 Irish general election. This study aims to establish the existence of different voting profiles in the electorate and to establish how voters align themselves with these profiles. This investigation will thus provide an enhanced insight into the actual voting behaviors exhibited in this electorate.

The Metropolis-within-Gibbs sampler, as outlined in Section 1.4.1, was run over 50000 iterations with a burn-in period of 10000 iterations. The model was fitted with $K = 1, 2, \ldots, 7$ voting profiles in order to establish the appropriate number of profiles to adequately model the data.

For each value of K, the DIC value was computed and these are shown in Figure 1.1.

The plot shows a sharply decreasing trend when K increases from 1 to 3 and the DIC values decrease slightly thereafter. Consequently, the fitted models for $K \geq 3$ were examined and it was determined that the $K = 3$ model was most appropriate.

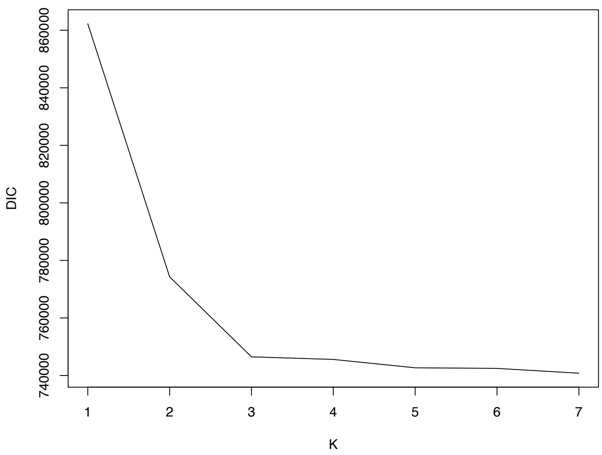


FIGURE 1.1
Values of the DIC for the mixed membership model for rank data fitted to the 2002 Dublin North constituency data over different values of the number of voting profiles K.

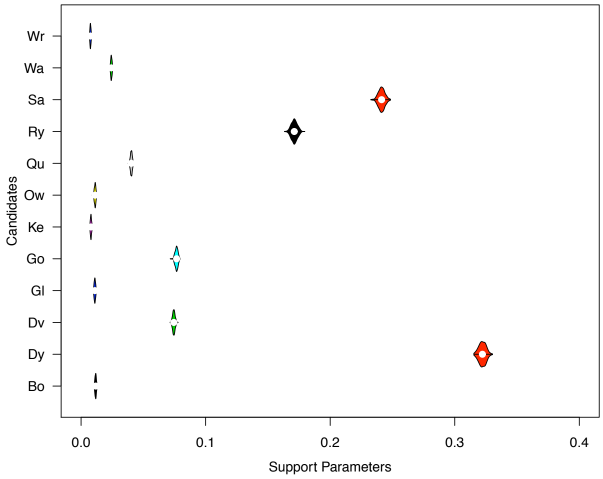
1.5.1 Support for the candidates

The marginal posterior density of the support parameters for each candidate within the three voting profiles are illustrated in Figure 1.2; a violin plot [26, 1] is used to show these marginal posterior densities.

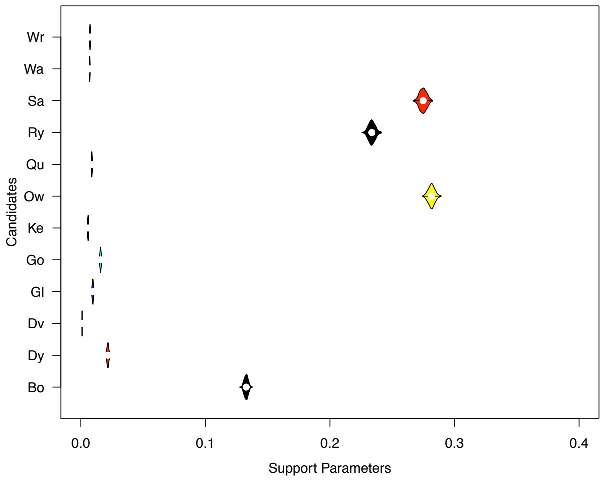
The three voting profiles have distinct and intuitive interpretations within the context of the 2002 Irish general election. The four elected candidates have high support in at least one of the voting profiles and some other prominent candidates also have high support.

Voting profile 1: non-mainstream opposition and protest voters.
(Figure 1.2(a)) The posterior mean support parameter estimate for the candidates in this voting profile suggest that this voting profile shows strong support for the non-mainstream opposition parties and single issue/protest candidates. Clare Daly (Socialist Party) has the largest support and she would be characterized as a major candidate in the non-mainstream opposition in Ireland; despite having such high support in this voting profile she failed to get elected. Trevor Sargent (Green Party) was leader of the Green Party at the time of the election and the 2002 election saw the party increase its number of seats in the Dáil from two to six seats thus moving them towards the mainstream opposition. Seán Ryan was a Labour party candidate, the Labour party has a diverse range of support within the Irish electorate so it could be considered to be a mainstream party but it would also have appeal to voters who don't support other mainstream parties. Interestingly, candidates that received very few first preference votes (eg. Eamon Quinn and David Walshe) have appreciable support in this voting profile.

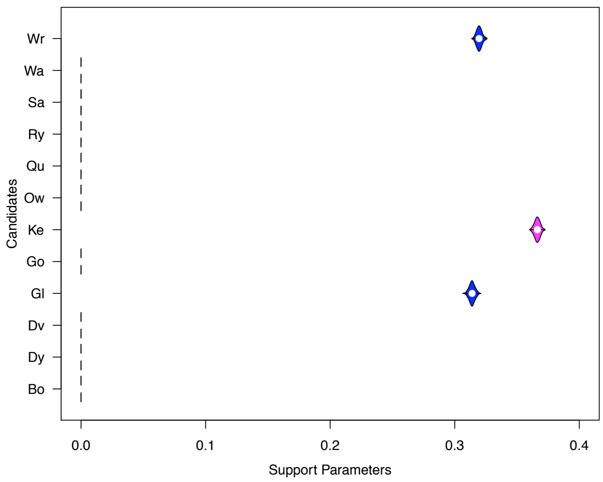
Voting profile 2: mainstream opposition voters.
(Figure 1.2(b)) The support parameters for Trevor Sargent (Green Party), Seán Ryan



(a) Voting profile 1. Non-mainstream opposition



(b) Voting profile 2. Mainstream opposition



(c) Voting profile 3. Fianna Fáil

FIGURE 1.2
Violin plots of the posterior samples for the support parameters. The plot shows the marginal posterior density for each support parameter, for each of the twelve candidates and the three voting profiles. The abbreviation used for each candidate's name is given in Table 1.1.

(Labour), Nora Owen (Fine Gael) and Cathal Boland (Fine Gael) are all large relative to the other candidates. Fine Gael were the largest opposition party before the election and their support here suggests that this voting profile shows support for the mainstream opposition parties. Labour were the second largest opposition party and traditionally Labour and Fine Gael have formed coalition governments, so they share much support amongst the voters. The 2002 election saw the Green party move towards becoming a mainstream opposition party and this is reflected in this voting profile too. Prior to the election, there was some discussion in the print media about Fine Gael, Labour and the Green party forming a coalition government if they gained enough seats, but this did not happen.

Voting profile 3: Fianna Fáil voters.
(Figure 1.2(c)) The posterior mean support parameter estimates for the candidates in this profile reveal that this profile only has appreciable support for the three Fianna Fáil candidates. All other candidates have very low support.

The division of the voters into three profiles provides a systematic method for decomposing the electorate into a small number of profiles. The relevance of the revealed profiles is supported by the exploratory analysis of these data in [32]. Interestingly, the division of candidates amongst the profiles corresponds very closely to the hierarchical decomposition of the candidates and parties in Dublin North as found in [27, 28].

1.5.2 Mixed membership parameters for the electorate

The unique feature of the mixed membership model is that the partial memberships of the voting profiles for each voter are inferred directly when estimating the model. The entropy [45] of each voter's mixed membership vector measures the degree to which they exhibit mixed membership across voting profiles. In fact, the exponential of the entropy can be seen as the effective number of profiles [6, 52] which are required to model voter i's preferences. Figure 1.3 shows a histogram of the exponentiated entropy values for the Dublin North voters and these show that there is significant evidence of mixed membership for the voters with many voters being effectively members of two or more of the profiles.

The voter with the lowest effective number of profiles has a membership vector $\underline{\pi}_i = (0.068, 0.885, 0.047)$ and they recorded the vote $\underline{x}_i$ =(Boland, Owen, Sargent, Ryan, Goulding, Quinn, Walsh, Daly, Glennon, Wright, Kennedy, Davis). Since their highest preference choices all have high support in voting profile 2, it is clear why they have particularly high membership to this profile and low membership to other profiles. The voter with the highest effective number of profiles has a membership vector $\underline{\pi}_i = (0.333, 0.336, 0.331)$ and they recorded the vote $\underline{x}_i$ =(Goulding, Daly, Ryan, Boland, Owen, Glennon, Wright, Kennedy). In this case, the voter's highest preference votes have high support in different profiles, so the mixed membership model suggests that all three profiles are needed to model their preferences.

We can further explore the mixed membership vectors by dividing the voters into groups by assigning each voter to the voting profile for which they have the highest membership score. We construct a kernel density estimate of the mixed membership parameter for each voting profile, for each of the groups of voters (Figure 1.4). Clearly, a significant proportion of the voters who have strongest affiliation to voting profiles 1 and 2 also have strong affiliation to at least one other profile. In contrast, voters who have strongest affiliation to voting profile 3 tend to have very little affiliation to the other voting profiles. This suggests that voting profiles 1 and 2 are closer and thus voters exhibit more mixed membership between these two profiles. This makes intuitive sense within the context of the 2002 Irish

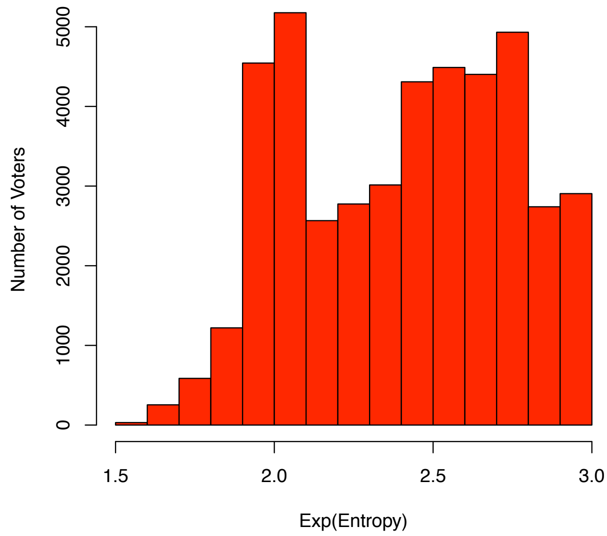


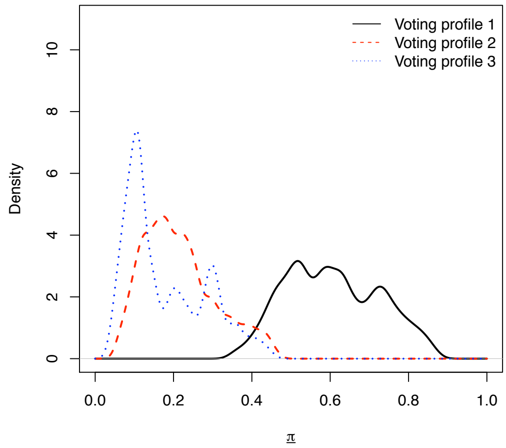
FIGURE 1.3
A histogram of the exponential of the entropy values for each voter's mixed membership parameter. The values shown give an "effective number of profiles" needed to model each voter.

general election as voting profile 3 represents the current government party, with profiles 1 and 2 representing two different types of opposition.

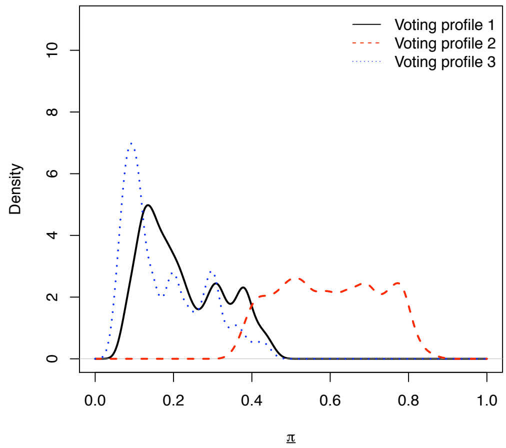
1.5.3 Posterior predictive model checks

Posterior predictive simulation [20] was employed to assess model fit. Subsequent to a burn-in period of 10000 iterations, 40000 samples thinned every 100th iteration were drawn from the posterior distribution $\mathbf{P}\{\boldsymbol{\pi}, \mathbf{p}, \mathbf{z}|\mathbf{x}\}$, giving $R = 400$ sets of parameters simulated from the posterior. A predictive election data set $\mathbf{x}^r$ was then simulated from the mixed membership model for rank data, given each of the $r = 1, \ldots, R$ draws of the parameters from the posterior distribution. Due to the discrete and structured nature of the data, it is difficult to fully assess model fit, so first order summaries were used. For the simulated votes, the number of first preference votes obtained by the twelve candidates was recorded. Figure 1.5 illustrates the number of first preferences received by each candidate in each simulated posterior predictive data set, and in the Dublin North voting data.

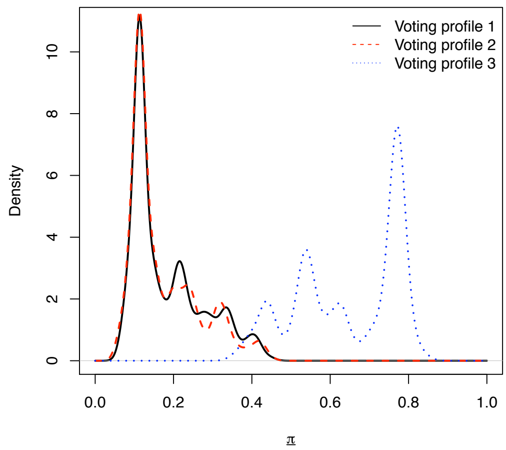
The posited model appears to capture the main structure of the data but there is some discrepancy between the observed and the simulated values. The discrepancy can be explained by the fact that the support parameters $\mathbf{p}$ are used to model the probability of candidate selection at all preference levels and thus the posterior estimates for these parameters depend on all preference levels rather than just first preferences. So, this may lead to a slight under or over estimation of the number of first preference selections for a candidate.



(a) Members of profile 1



(b) Members of profile 2



(c) Members of profile 3

FIGURE 1.4
Kernel density estimates of the membership parameters for those voters most likely to be characterized by each profile.

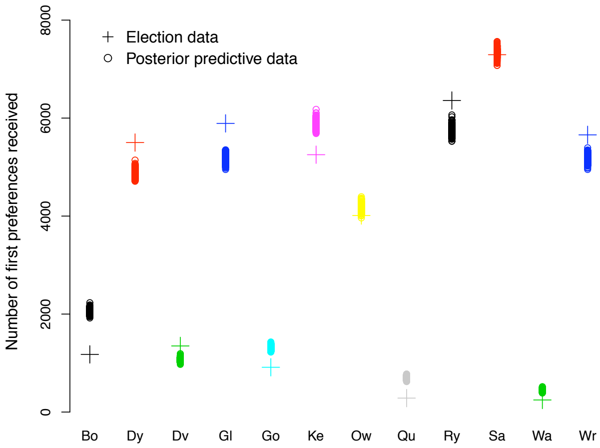


FIGURE 1.5
This plot shows the posterior predictive counts for each candidate in the Dublin North constituency. Each circle indicates the number of first preference votes received by the twelve candidates in each of 400 simulated posterior predictive data sets. The crosses indicate the number of first preferences received by each candidate in the actual voting data.

1.6 Conclusions

A mixed membership model for rank data has been described and applied to the analysis of a large election data set. It has been shown that in the context of analyzing rank response data, the model provides scope to examine a population for the presence of preference profiles, to estimate the characteristics of these profiles and to investigate the mixed membership of population members to the profiles on a case by case basis. The loss of information which may result from a hard clustering of the data is avoided by providing a soft clustering of the population. In addition, the mixed membership model provides a parsimonious description of the population because complex preference patterns can be captured using the mixed membership machinery.

The method provides an alternative modeling framework to the many mixture modeling approaches for rank data [49, 41, 21, 5, 22, 40]. In particular, [22] developed a finite mixture of Plackett-Luce models for modeling PR-STV data which provides a competing modeling framework. However, when studying large voting data sets with diverse candidates, a large number of mixture components are needed to appropriately model the data. In contrast, the mixed membership model can model voting in such elections with many fewer profiles.

The model described herein can be fitted in a Bayesian paradigm using an efficient Markov chain Monte Carlo scheme. The method is able to explore the posterior efficiently because the proposal distributions developed for sampling the support parameters, which don't have a closed form conditional posterior, are accurate approximations of the parameter conditional posterior distributions. Recently, [8] developed a Gibbs sampling method for the Plackett-Luce model and this could be adapted to fitting the mixed membership model outlined herein, thus improving the accuracy of model inference. An alternative method for fitting such models would be to use variational Bayesian (VB) methods or expectation propagation (EP); [51] developed an online VB algorithm and [25] developed an EP algorithm for a single Plackett-Luce model and there is potential to extend these methods to the mixed membership model herein.

The mixed membership model for rank data could be developed in several directions. In terms of the application in this article further model accuracy could be attained by imposing a hierarchical framework — a hyperprior could be introduced for the Dirichlet parameters $\underline{\alpha}$ and $\underline{\beta}$ of the mixed membership and support parameter priors respectively; such hierarchical priors are employed in [44] and [14]. The issue of model choice for mixed membership models is still problematic [30]. The combination of the use of DIC [47] and posterior predictive model checks [20] provided a suitable method in this application, but there were a number of competing models that achieved similar fit. Thus, there remains the need for more automatic model choice methods.

Recently, a number of models have been developed that capture underlying group structure for rank data when concomitant information for the voters is also available [23, 17, 33, 34, 35]. It would be worthwhile to extend the mixed membership modeling framework for rank data to include such concomitant information. Such a modeling extension would help explain the structure revealed by the mixed membership model for ranked data.

Appendix A: Data sources

The 2002 Dublin North constituency voting data was made available by the Dublin County Returning Officer. The data are available from the authors on request.

Bibliography

[1] Daniel Adler. *vioplot: Violin plot*, 2005. R package version 0.2.

[2] E. M. Airoldi, E. A. Erosheva, S. E. Fienberg, C. Joutard, T. Love, and S. Shringarpure. Re-conceptualizing the classification of PNAS articles. *Proceedings of the National Academy of Sciences*, 107(49):20899–20904, 2010.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[4] Ralph A. Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39:324–345, 1952.

[5] Ludwig M. Busse, Peter Orbanz, and Joachim M. Buhmann. Cluster analysis of heterogeneous rank data. In *Proceedings of the 24th international conference on Machine learning*, ICML '07, pages 113–120, New York, NY, USA, 2007. ACM.

[6] L. L. Campbell. Exponential entropy as a measure of extent of a distribution. *Probability Theory And Related Fields*, 5(3):217–225, 1966.

[7] Bradley P. Carlin and Thomas A. Louis. *Bayes and empirical Bayes methods for data analysis.* Chapman & Hall, New York, 2nd edition, 2000.

[8] François Caron and Arnaud Doucet. Efficient Bayesian inference for generalized Bradley-Terry models. *Journal of Computational & Graphical Statistics*, 21:174–196, 2012.

[9] Gilles Celeux, Merrilee Hurn, and Christian P. Robert. Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970, 2000.

[10] E. A. Erosheva. Latent class representation of the Grade of Membership Model. Technical Report 492, Department of Statistics, University of Washington, 2006.

[11] E. A. Erosheva, S. E. Fienberg, and C. Joutard. Describing disability through individual-level mixture models for multivariate binary data. *The Annals of Applied Statistics*, 1(2):502–537, 2007.

[12] E. A. Erosheva, S. E. Fienberg, and J. Lafferty. Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences*, 101(Suppl.1):5220–5227, 2004.

[13] E.A. Erosheva. *Grade of membership and latent structure models with application to disability survey data.* PhD thesis, Department of Statistics, Carnegie Mellon University, 2002.

[14] Elena A. Erosheva. Bayesian Estimation of the Grade of Membership Model. In J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West, editors, *Bayesian Statistics, 7*, pages 501 – 510. Oxford University Press, Oxford, 2003.

[15] David M. Farrell. *Electoral Systems: A Comparative Introduction*. St. Martin's Press, New York, 2001.

[16] Chris Fraley and Adrian E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631, 2002.

[17] Brian Francis, Regina Dittrich, and Reinhold Hatzinger. Modeling heterogeneity in ranked responses by nonparametric maximum likelihood: How do Europeans get their scientific knowledge? *Annals of Applied Statistics*, 4(4):2181–2202, 2010.

[18] Michael Gallagher, Michael Marsh, and Paul Mitchell. *How Ireland Voted 2002*. Palgrave Macmillian, Basingtoke, 2003.

[19] Michael Gallagher and Paul Mitchell. *The Politics of Electoral Systems*. Oxford University Press, Oxford, 2005.

[20] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov chain Monte Carlo in practice*. Chapman & Hall, London, 1996.

[21] I. C. Gormley and T. B. Murphy. Analysis of Irish third-level college applications data. *Journal of the Royal Statistical Society, Series A*, 169(2):361—379, 2006.

[22] I. C. Gormley and T. B. Murphy. Exploring voting blocs within the Irish electorate: A mixture modeling approach. *Journal of the American Statistical Association*, 103(483):1014–1027, 2008.

[23] I. C. Gormley and T. B. Murphy. A mixture of experts model for rank data with applications in election studies. *The Annals of Applied Statistics*, 2(4):1452–1477, 2008.

[24] I.C. Gormley and T.B. Murphy. A grade of membership model for rank data. *Bayesian Analysis*, 4(2):265–296, 2009.

[25] John Guiver and Edward Snelson. Bayesian inference for Plackett-Luce ranking models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 377–384, New York, NY, USA, 2009. ACM.

[26] Jerry L. Hintze and Ray D. Nelson. Violin plots: A box plot-density trace synergism. *The American Statistician*, 52:181–184, 1998.

[27] Jonathan Huang. *Probabilistic Reasoning and Learning on Permutations: Exploiting structural decompositions of the symmetric group*. PhD thesis, Carnegie Mellon University, 2011.

[28] Jonathan Huang and Carlos Guestrin. Uncovering the riffled independence structure of rankings. *Electronic Journal of Statistics*, 6:199–230, 2012.

[29] A. Jasra, C. C. Holmes, and D. A. Stephens. Markov chain Monte Carlo Methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, 20:50–67, 2005.

[30] C.J. Joutard, E.M. Airoldi, S.E. Fienberg, and T.M. Love. Discovery of latent patterns with hierarchical Bayesian mixed-membership models and the issue of model choice. In P. Poncelet, F. Masseglia, and M. Teisseire, editors, *Data Mining Patterns: New Methods and Applications*. IGI Global, Pennsylvania, 2008.

[31] Fiachra Kennedy. The 2002 general election in Ireland. *Irish Political Studies*, 17(2):95–106, 2002.

[32] Michael Laver. Analysing structure of party preference in electronic voting data. *Party Politics*, 10:521–541, 2004.

[33] Paul H. Lee and Philip L. H. Yu. Distance-based tree models for ranking data. *Computational Statistics & Data Analysis*, 54(6):1672–1682, 2010.

[34] Paul H. Lee and Philip L. H. Yu. Mixtures of weighted distance-based models for ranking data with applications in political studies. *Computational Statistics & Data Analysis*, 56(8):2486–2500, 2012.

[35] Jianbo Li, Minggao Gu, and Tao Hu. General partially linear varying-coefficient transformation models for ranking data. *Journal of Applied Statistics*, page To appear, 2012.

[36] C. L. Mallows. Non-null ranking models. I. *Biometrika*, 44:114–130, 1957.

[37] John I. Marden. *Analyzing and modeling rank data*. Chapman & Hall, London, 1995.

[38] Michael Marsh. The Irish general election of 2002: A new hegemony for Fianna Fail? *British Elections & Parties Review*, 13:17–28, 2003.

[39] Geoffrey J. McLachlan and Kaye E. Basford. *Mixture models: Inference and applications to clustering*. Marcel Dekker Inc., New York, 1988.

[40] Marina Meila and Harr Chen. Dirichlet process mixtures of generalized Mallows models. In Peter Grünwald and Peter Spirtes, editors, *UAI*, pages 358–367. AUAI Press, 2010.

[41] Thomas Brendan Murphy and Donal Martin. Mixtures of distance-based models for ranking data. *Computational Statistics and Data Analysis*, 41(3–4):645–655, 2003.

[42] Anthony O'Hagan and Jonathon Forster. *Kendall's Advanced Theory of Statistics: Volume 2B Bayesian Inference*. Arnold, London, UK, second edition, 2004.

[43] R. L. Plackett. The analysis of permutations. *Applied Statistics*, 24(2):193–202, 1975.

[44] Johnathon K. Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959, 2000.

[45] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[46] Richard Sinnott. The electoral system. In John Coakley and Michael Gallagher, editors, *Politics in the Republic of Ireland*, pages 99–126. Routledge & PSAI Press, London, 3rd edition, 1999.

[47] David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64(4):583–639, 2002.

[48] Matthew Stephens. Dealing with label-switching in mixture models. *Journal of the Royal Statistical Society, Series B*, 62(4):795–810, 2000.

[49] Hal S. Stern. Probability models on rankings and the electoral process. In Michael A. Fligner and Joseph S. Verducci, editors, *Probability Models and Statistical Analyses For Ranking Data*, pages 173–195. Springer-Verlag, New York, 1993.

[50] Liam Weeks. The Irish parliamentary election, 2002. *Representation*, 39(3):215–225, 2002.

[51] Ruby C. Weng and Chih-Jen Lin. A Bayesian approximation method for online ranking. *Journal of Machine Learning Research*, 12:267–300, 2011.

[52] Arthur White, Jeffrey Chan, Conor Hayes, and Thomas Brendan Murphy. Mixed membership models for exploring user roles in online fora. In *Proceedings of the Sixth International Conference on Weblogs and Social Media*, 2012.