16

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
16.1 Introduction .................................. 297
16.2 The 2002 Irish General Election ..................... 298
16.3 Model Specification ................................ 300
  16.3.1 The Plackett-Luce Model for Rank Data ............. 300
  16.3.2 The Mixed Membership Model for Rank Data .......... 302
  A Latent Class Representation of the Mixed Membership Model ... 303
  16.3.3 Prior and Posterior Distributions ................. 304
16.4 Model Inference .................................... 305
  16.4.1 Parameter Estimation ............................ 305
  16.4.2 Model Selection ................................. 306
16.5 Application to the 2002 Irish General Election .......... 307
  16.5.1 Support for the Candidates ...................... 307
  16.5.2 Mixed Membership Parameters for the Electorate .... 310
  16.5.3 Posterior Predictive Model Checks .............. 312
16.6 Conclusions ...................................... 312
Appendix : Data Sources ................................. 315

A mixed membership model is an individual-level mixture model where individuals have partial membership of the profiles 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 dataset 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.
16.1 Introduction

Mixture models are a well established tool for model-based clustering of data (McLachlan and Basford, 1988; Fraley and Raftery, 2002). 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 to have partial membership in each of the profiles 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 Erosheva (2002) and Blei et al. (2003), and this model has been used in a number of applications including Erosheva et al. (2004; 2007) and Airoldi et al. (2010), 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 Gormley and Murphy (2009) 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 16.2 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 is discussed in Section 16.3.1. The specification of the mixed membership model for rank data follows in Section 16.3.2. Estimation of the mixed membership model for rank data is outlined in Section 16.4.1. The question of model choice is addressed in Section 16.4.2. The application of the mixed membership model for rank data to 2002 Irish general election data is given in Section 16.5. The article concludes in Section 16.6 with a discussion of the methodology.
16.2 The 2002 Irish General Election

Dáil Éireann is the main parliament in the Republic of Ireland; it has 166 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 May 17, 2002, a general election was held to elect the 29th Dáil; candidates ran in 42 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 Ceann Comhairle is the position of Speaker of the House in Dáil Éireann. The Ceann Comhairle from the previous parliament is automatically re-elected in their constituency and thus the number of candidates elected through the general election in that constituency is reduced by one. The outgoing government consisted of a Fianna Fáil and Progressive Democrat coalition with Fianna Fáil having 77 seats and the Progressive Democrats having 4 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, this time with a majority and holding 81 and 8 seats, respectively. This was the first time that a government had been re-elected in an Irish general election in 30 years. Extensive descriptions of the 2002 election are provided by Kennedy (2002), Weeks (2002), Gallagher et al. (2003), and Marsh (2003).

In 2002, the Dublin North constituency consisted of an electorate of 72,353 with four TDs to be elected from this constituency. A total of 43,942 people voted and twelve candidates ran for election: Fianna Fáil, the largest political party at the time, ran three candidates; Fine Gael, the largest opposition party, ran two candidates; the Labour, Green, and Sinn Féin parties ran one candidate each, and smaller 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.

The votes in the election were totaled 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 Sinnott (1999) and a good overall comparison of different voting systems is given by Farrell (2001) and Gallagher and Mitchell (2005).
Details of the counting and transfer of votes in the Dublin North constituency are shown in Table 16.1. The total valid poll was 43,942, so the number of votes required to guarantee election (called the droop quota) was 8,789 votes. In the first count, the number of first preferences for each candidate was counted. If no candidate exceeded the droop quota, then the lowest candidates were eliminated and their votes were distributed using the next available preferences on their ballots; i.e., a vote was transferred to the next preferred candidate on the ballot who had not been eliminated or already elected, if no such candidate existed then the vote was considered to be non transferable. If a candidate was elected by exceeding the droop quota, then their surplus votes (the amount by which they exceed the droop quota) were distributed using the next available preferences on these surplus votes; the surplus votes to be transferred were sampled from the set of votes that brought the candidate over the droop quota. The procedure of eliminating low candidates and distributing surpluses continued until either four candidates exceeded the droop quota or only four candidates remained.

For example, Trevor Sargent was the first candidate to be elected; he was elected in round 6 of the count because he exceeded the droop quota on the basis of the 7,294 first preference votes and 2,491 votes that he received through transfers in rounds 1 to 5 of the count. Because he received 997 votes in excess of the droop quota, these excess votes were transferred in round 7; the 997 votes that were distributed were sampled from the 1,667 that he received in round 5, because these votes brought his total over the droop quota.

By the end of the vote count in Dublin North, two candidates reached the droop 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.

### 16.3 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 (e.g., Gormley and Murphy, 2008a) 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.
Mixed Membership Models for Rank Data

TABLE 16.1
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.

<table>
<thead>
<tr>
<th>Candidate (Abbreviation)</th>
<th>Party (Abbreviation)</th>
<th>Round of Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1  2  3  4  5  6  7  8</td>
</tr>
<tr>
<td>Trevor Sargent*</td>
<td>Green</td>
<td>7294 7380 7678 7818 8118 9785 8789 8789</td>
</tr>
<tr>
<td>(Sa)</td>
<td></td>
<td>+86 +298 +140 +300 +1667 -996</td>
</tr>
<tr>
<td>Seán Ryan*</td>
<td>Labour</td>
<td>6359 6407 6535 6665 6847 8578 9128 9128</td>
</tr>
<tr>
<td>(Ry)</td>
<td></td>
<td>+48 +128 +130 +182 +1731 +550</td>
</tr>
<tr>
<td>Jim Glennon</td>
<td>Fianna Fáil (Gl)</td>
<td>5892 5945 6028 6152 6294 6511 6598 8640</td>
</tr>
<tr>
<td>(Gl)</td>
<td></td>
<td>+53 +83 +124 +142 +217 +85 +2044</td>
</tr>
<tr>
<td>G V Wright*</td>
<td>Fianna Fáil (Wr)</td>
<td>5658 5707 5739 5777 5868 6139 6249 8617</td>
</tr>
<tr>
<td>(Wr)</td>
<td></td>
<td>+49 +32 +38 +91 +271 +110 +2368</td>
</tr>
<tr>
<td>Clare Daly</td>
<td>Socialist (Dy)</td>
<td>5501 5551 5730 5796 6244 6590 6772 7523</td>
</tr>
<tr>
<td>(Dy)</td>
<td></td>
<td>+53 +179 +66 +448 +346 +182 +751</td>
</tr>
<tr>
<td>Michael Kennedy</td>
<td>Fianna Fáil (Ke)</td>
<td>5253 5309 5368 5422 5532 5732 5801</td>
</tr>
<tr>
<td>(Ke)</td>
<td></td>
<td>+56 +59 +54 +110 +200 +69 -5801</td>
</tr>
<tr>
<td>Nora Owen*</td>
<td>Fine Gael (Ow)</td>
<td>4012 4030 4132 4720 4763</td>
</tr>
<tr>
<td>(Ow)</td>
<td></td>
<td>+18 +102 +588 +43 -4763</td>
</tr>
<tr>
<td>Mick Davis</td>
<td>Sinn Féin (Dv)</td>
<td>1350 1382 1424 1440</td>
</tr>
<tr>
<td>(Dv)</td>
<td></td>
<td>+32 +42 +16 -1440</td>
</tr>
<tr>
<td>Cathal Boland</td>
<td>Fine Gael (Bo)</td>
<td>1177 1189 1216</td>
</tr>
<tr>
<td>(Bo)</td>
<td></td>
<td>+12 +27 -1216</td>
</tr>
<tr>
<td>Ciarán Goulding</td>
<td>Independents (Go)</td>
<td>914 1009</td>
</tr>
<tr>
<td>(Go)</td>
<td>Health Alliance</td>
<td>+95 -1009</td>
</tr>
<tr>
<td>Eamon Quinn</td>
<td>Independent (Qu)</td>
<td>285</td>
</tr>
<tr>
<td>(Qu)</td>
<td></td>
<td>-285</td>
</tr>
<tr>
<td>David Walsh</td>
<td>Christian (Wa)</td>
<td>247</td>
</tr>
<tr>
<td>(Wa)</td>
<td>Solidarity Party</td>
<td>-247</td>
</tr>
<tr>
<td>Non Transferable</td>
<td></td>
<td>33 92 152 276 607 607 1245</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>43,942</td>
</tr>
</tbody>
</table>

16.3.1 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 (Bradley and Terry, 1952; Mallows, 1957; Plackett, 1975), and these are reviewed in Marden (1995). In this study the Plackett-Luce model (Plackett, 1975) is utilized to model the rank nature of the data.

The Plackett-Luce model is parameterized by a ‘support’ parameter

\[ p = (p_1, p_2, \ldots, p_N), \]

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 model assumes that probability of candidate \(j\) being given a lower than first preference is proportional to their support parameter \(p_j\) but conditional on a smaller number of candidates being available for selection at lower preferences. Hence, at preference levels lower than the first the probabilities are re-normalized to provide valid probability values. Further, it can be shown that the Plackett-Luce model has a random utility choice model interpretation (Chapman and Staelin, 1982).

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

\[
P\{x_i | p\} = \prod_{t=1}^{n_i} \frac{p_{c(i,t)}}{p_{c(i,t)} + p_{c(i,t+1)} + \cdots + p_{c(i,N)}}
\]

\[
= \prod_{t=1}^{n_i} \frac{p_{c(i,t)}}{\sum_{s=t}^{N} p_{c(i,s)}} = \prod_{t=1}^{n_i} \frac{1}{q_{it}}, \quad (16.1)
\]

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 (16.1) sums to 1 over all \(n_i!\) possible permutations of the candidates ranked in the vote \(x_i\).

### 16.3.2 The Mixed Membership Model for Rank Data

Mixed membership models allow every individual in a population to have partial membership in 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 Gormley and Murphy (2009).

Under the mixed membership model, each voter \(i = 1, \ldots, M\) has an associated \textit{mixed membership parameter} \(\pi_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{iK})\) which is a direct parameter of the model. The mixed membership parameter \(\pi_i\) 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 \(K \subset \{1, 2, \ldots, K\}\), then \(\pi_{ij} > 0\) for \(j \in K\) and \(\pi_{ij} = 0\) for \(j \not\in K\).

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 voter’s 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 \(\pi_i\) and the profile specific support parameters \(p = (p_1, p_2, \ldots, p_K)\),
Mixed Membership Models for Rank Data

is given as

\[ P\{c(i, t) = j \mid \pi, p\} = \sum_{k=1}^{K} \pi_{ik} \left[ \frac{p_{kj}}{\sum_{k=1}^{N} p_{kc(i,s)}} \right] . \] (16.2)

Additionally, local independence is then assumed between each preference level \( t \), given the mixed membership parameters. Thus, the conditional probability of ranking \( x \) given membership parameter \( \pi \) and support parameters \( p \) is

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

and the likelihood function based on the data \( x = (x_1, x_2, \ldots, x_M) \) is therefore

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

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 (Stern, 1993; Murphy and Martin, 2003; Gormley and Murphy, 2006; Busse et al., 2007; Gormley and Murphy, 2008a;b). Modeling rank data in this manner provides a deeper insight into 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 16.3.2).

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 Erosheva (2006); 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 \( \pi_i \) for voter \( i \).

For each voter \( i \), we impute binary latent vectors \( \tilde{z}_{it} = (\tilde{z}_{it1}, \ldots, \tilde{z}_{itK}) \) for \( t = 1, \ldots, n_i \), where \( \tilde{z}_{it} \sim \text{Multinomial}(1, \pi_i) \). The value of \( \tilde{z}_{it} \) 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 \( x \) and the binary latent variables \( z \) is therefore of the
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 \(t\).

### 16.3.3 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\((\alpha)\) distribution and that the support parameters follow a Dirichlet\((\beta)\) distribution, i.e.,

\[
\pi_i \sim \text{Dirichlet}\{\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_K)\}
\]

\[
p_k \sim \text{Dirichlet}\{\beta = (\beta_1, \beta_2, \ldots, \beta_N)\}.
\]

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 Gormley and Murphy (2009). For even moderate sized datasets it was found that the posterior inferences were not heavily influenced by the prior specification. In Gormley and Murphy (2009), the sensitivity of the choice of prior model and hyperparameters was considered. In practice, the prior parameters are fixed as \(\alpha = (0.5, \ldots, 0.5)\) and \(\beta = (0.5, \ldots, 0.5)\), which is the Jeffreys prior for the multinomial distribution (e.g., O’Hagan and Forster, 2004). These priors have positive mass near the corners of the parameter simplex and thus the posterior distributions of the parameters can have high probability in these regions. However, the choice of parameters also avoids the posterior concentrating exactly on the corners of the simplex.

In principle, the prior hyperparameters could be estimated as part of the inference procedure rather than fixed as done here, but this greatly increases the computational burden of model fitting and inference.

Given these prior distributions and the augmented data likelihood function (16.3) from the mixed membership model for rank data, the posterior distribution based on the data is:
Mixed Membership Models for Rank Data

This posterior distribution differs from the posterior distribution in the case of the original mixed membership model (Erosheva, 2002; 2003) 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.

16.4 Model Inference

16.4.1 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 (MCMC) 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, it is not.

The full conditional distributions of the latent variables $z_{it}$ and the mixed membership parameters $\pi_i$ are readily available. In particular,

$$z_{it} \sim \text{Multinomial} \left\{ 1, \left( \frac{\pi_{1i}q_{1it}}{\sum_{k'=1}^{K} \pi_{ik'}q_{k'it}}, \frac{\pi_{2i}q_{2it}}{\sum_{k'=1}^{K} \pi_{ik'}q_{k'it}}, \ldots, \frac{\pi_{Ki}q_{Kit}}{\sum_{k'=1}^{K} \pi_{ik'}q_{k'it}} \right) \right\},$$

where $q_{kit}$ is defined as in (16.1) for $k = 1, 2, \ldots, K$, $i = 1, \ldots, M$, $t = 1, \ldots, n_i$ and

$$\pi_i \sim \text{Dirichlet} \left( \alpha_1 + \sum_{t=1}^{n_i} z_{1it}, \ldots, \alpha_K + \sum_{t=1}^{n_i} z_{Kit} \right),$$

for $i = 1, \ldots, M$.

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

$$P\{p_k|x, z\} \propto \left[ \prod_{i=1}^{M} \prod_{t=1}^{n_i} \pi_{ik} \frac{P_k(c_{i,t})}{\sum_{s=t}^{N} P_k(c_{i,s})} \right]^{z_{ikt}} \left[ \prod_{j=1}^{N} p_{kj}^{\beta_j-1} \right]. \quad (16.4)$$
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 (Carlin and Louis, 2000) 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.

We start to construct a proposal distribution by examining the logarithm of the full conditional of the support parameter \( p_k \) (16.4) which is of the form

\[
\log P\{p_k | \pi, x, z\} \propto \sum_{i=1}^{M} \sum_{t=1}^{n_i} \sum_{t_{i,t} \in k} \left( \log \frac{p_k}{\beta_j} - \log \sum_{s=t}^{N} \frac{p_k(s)}{\beta_j} \right) + \sum_{j=1}^{N} (\beta_j - 1) \log \frac{1}{p_k(s)}.
\]

The function \(-\log(\cdot)\) is a convex function and thus the term \(-\log \sum_{s=t}^{N} p_k(s)\) can be approximated (in fact lower bounded) by a hyperplane that is tangent to the function at the currently sampled value of \( p_k \). 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 Stephens (2000), Celeux et al. (2000), and Jasra et al. (2005) need to be considered to avoid this issue. The online relabeling algorithm of Stephens (2000) was found to be an effective method for handling this issue; this algorithm implements relabeling as the MCMC algorithm progresses rather than as a post-processing step.

Full details of the Metropolis-within-Gibbs algorithm for fitting this model are given in Gormley and Murphy (2009).

### 16.4.2 Model Selection

Another feature of the mixed membership model is the need to infer the model dimensionality, i.e., the number of voting profiles \( K \) needed to appropriately model the electorate. Within the Bayesian paradigm, the natural approach would be to base
inference on the posterior distribution of $K$ given the data $x$, $P\{K|x\}$. 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 Joutard et al. (2008).

In this application of the mixed membership model for rank data, the Deviance Information Criterion (DIC), introduced by Spiegelhalter et al. (2002), 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 $x$ and parameters $\theta$ the DIC is

$$DIC = D(\theta) + p_D,$$

where $D(\theta) = -2 \log[P(x|\theta)] + 2 \log[h(x)]$ is the Bayesian deviance and $h(x)$ is a function of the data only. The effective number of parameters is defined as $p_D = D(\theta) - D(\bar{\theta})$. The criterion has an approximate decision theoretic justification. In any case, models with small DIC values are preferable to models with large DIC values. The choice of $\theta$ in the calculation of DIC is important, and we use $\theta = (\pi, p)$ because these are the primary model parameters of interest.

### 16.5 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 16.4.1, was run over 50,000 iterations with a burn-in period of 10,000 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 (shown in Figure 16.1). 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 because the models with $K > 3$ included extra extreme profiles that didn’t differ greatly from those in the model with $K = 3$.  


FIGURE 16.1
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 $K$.

16.5.1 Support for the Candidates
The marginal posterior density of the support parameters for each candidate within the three voting profiles are illustrated in Figure 16.2; a violin plot (Hintze and Nelson, 1998; Adler, 2005) is used to show these marginal posterior densities. The violin plot combines a boxplot and a kernel density estimate; the length of the violin corresponds to the length of the box in a boxplot but the breadth of the violin shows a back-to-back plot of a kernel density estimate of the values. The marginal probabilities for the voting profiles are $(0.323, 0.324, 0.353)$, respectively.

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 16.2(a). The posterior mean support parameter estimates for the candidates in this voting profile suggest that a pure member of this voting profile should strongly support 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
Mixed Membership Models for Rank Data

FIGURE 16.2
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 16.1. The elected candidates are marked with an asterisk.

(a) Voting Profile 1: Non-mainstream opposition.

(b) Voting Profile 2: Mainstream opposition.

(c) Voting Profile 3: Fianna Fáil.
diverse range of support within the Irish electorate so it could be considered to be a mainstream party, but it would also have to appeal to voters who don’t support other mainstream parties. Interestingly, candidates that received very few first preference votes (e.g., Eamon Quinn and David Walshe) have appreciable support in this voting profile. The non-election of Claire Daly, despite having high support, can be explained by the fact that Trevor Sargent and Seán Ryan were only elected in the later counts (see Table 16.1), so Claire Daly didn’t have the opportunity to receive transfers from voters who gave the other candidates higher preferences than her.

Voting Profile 2: Mainstream opposition voters.
Figure 16.2(b). The support parameters for Trevor Sargent (Green Party), Seán Ryan (Labour), Nora Owen (Fine Gael), and Cathal Boland (Fine Gael) are all large relative to the other candidates. Fine Gael was the largest opposition party before the election and their support here suggests that this voting profile shows support for the mainstream opposition parties. Labour was 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; 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 16.2(c). The posterior mean support parameter estimates for the candidates in this profile reveal that only those voters with a high degree of profile membership should give strong support to 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 Laver (2004). 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 Huang (2011) and Huang and Guestrin (2012).

16.5.2 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 (Shannon, 1948) 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 (Campbell, 1966; White et al., 2012) which are required to model voter i’s preferences. Figure 16.3 shows a histogram of the exponentiated entropy values for the Dublin North
voters. These show that there is significant evidence of mixed membership for the voters with many being effectively members of two or more of the profiles.

![Figure 16.3](image_url)

**FIGURE 16.3**
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.

The voter with the lowest effective number of profiles has a membership vector \( \pi_i = (0.068, 0.885, 0.047) \) and they recorded the vote \( x_i = \) (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 \( \pi_i = (0.333, 0.336, 0.331) \) and they recorded the vote \( x_i = \) (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, assigning each voter to the voting profile for which they have the highest membership score (i.e., their modal profile membership). We construct a kernel density estimate of the mixed membership parameter for each voting profile for each of the groups of voters (Figure 16.4). Clearly, a significant proportion of the voters who have the strongest affiliation to Voting Profiles 1 and 2 also have a 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, thus voters exhibit more mixed membership between these two profiles. This makes intuitive sense within the context of the 2002
Irish general election as Voting Profile 3 represents the current government party, with Profiles 1 and 2 representing two different types of opposition.

16.5.3 Posterior Predictive Model Checks

Posterior predictive simulation (Gilks et al., 1996) was employed to assess model fit. Subsequent to a burn-in period of 10,000 iterations, 40,000 samples thinned every 100th iteration were drawn from the posterior distribution $P(\pi, p, z|x)$, giving $R = 400$ sets of parameters simulated from the posterior. A predictive election dataset $x^r$ was then simulated from the mixed membership model for the 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 16.5 illustrates the number of first preferences received by each candidate in each simulated posterior predictive dataset, 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 $p$ 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.

16.6 Conclusions

A mixed membership model for rank data has been described and applied to the analysis of a large election dataset. 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 particular, a hard clustering forces each voter to belong to one and only one cluster, so even if they are best characterized by a single cluster, any unusual aspects of their voting preferences are lost in the hard clustering. In contrast, the mixed membership model provides a parsimonious description of voting preferences 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 (Stern, 1993; Murphy and Martin, 2003; Gormley and Murphy, 2006; Busse et al., 2007; Gormley and Murphy, 2008a; Meilã and Chen, 2010). In particular, Gormley and Murphy (2008a) developed a finite mix-
FIGURE 16.4
Kernel density estimates of the membership parameters for those voters most likely to be characterized by each profile.
Number of first preferences received
0 2000 4000 6000 8000
Bo Dy Dv Gl Go Ke Ow Qu Ry Sa Wa Wr
Election data
Posterior predictive data

FIGURE 16.5
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 datasets. The crosses indicate the number of first preferences received by each candidate in the actual voting data.

ture of Plackett-Luce models for modeling PR-STV data which provides a modeling framework. However, when studying large voting datasets with diverse candidates, a large number of mixture components are needed to appropriately model the data. In contrast, the mixed membership model can represent 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, Caron and Doucet (2012) 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); Weng and Lin (2011) developed an online VB algorithm and Guiver and Snelson (2009) developed an EP algorithm for a single Plackett-Luce model; 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 chapter, further model accuracy could be attained by imposing a hierarchical framework—a hyperprior could be introduced for the Dirichlet parameters $\alpha$ and $\beta$ of the mixed membership and support parameter priors, respectively; such hierarchical priors are employed in Pritchard et al. (2000) and Erosheva (2003). The issue of model choice for mixed membership models is still problematic (Joutard et al., 2008). The combination of the use of DIC
Mixed Membership Models for Rank Data

(Spiegelhalter et al., 2002) and posterior predictive model checks (Gilks et al., 1996) provided a suitable method in this application, but there were different numbers of extreme profiles ($K$) 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 (Gormley and Murphy, 2008b; Francis et al., 2010; Lee and Yu, 2010; 2012; Li et al., 2012). 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 : 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.
References


