Fitting Mixed Membership Models using `mixedMem`

Y. Samuel Wang and Elena A. Erosheva

February 6, 2015

Abstract

This vignette is a tutorial for the `mixedMem` R package that allows users to fit and visualize results for a broad class of mixed membership models as specified by Erosheva et al. [2004]. Relevant data structures include multivariate outcome vectors. Mixed distribution types and replicate measurements are supported. Version 1.0 allows for binary, multinomial or Plackett-Luce rank distribution types. This tutorial provides an outline of functions available in the package and a step-by-step guide for fitting mixed membership models. As an example, we carry out a mixed membership analysis to identify latent ideology blocs using political survey data from the 1983 American National Election Survey Pilot.

1 Mixed Membership Modeling

1.1 Mixed Membership Background

Mixed membership models provide a useful framework for analyzing multivariate data from heterogeneous populations [Airoldi et al., 2014]. Similar to mixture models, mixed membership models assume that the overall population is comprised of several latent sub-populations or groups. Mixture models, however, assume that each individual within the population belongs to a single group, whereas mixed membership models allow for an individual to belong to multiple groups simultaneously and explicitly specify an individual's degree of membership in each group. (ELENA TO FILL IN)

Mixed membership models have been used to analyze to a wide variety of data types. For example, mixed membership analysis of text data [Blei et al., 2003, Erosheva et al., 2004] can illustrate the latent structure of topics/subjects within a body of documents; analysis of genotype sequences [Pritchard et al., 2000] can describe individuals sharing ancestry across various origin populations; analysis of survey data [Erosheva et al., 2007, Gross and Manrique-Vallier, 2014] can show how survey respondents belong to various sub-populations; and analysis of ranked votes [Gormley and Murphy, 2009] can characterize latent voting blocs and voting tendencies. The `mixedMem` package allows for fitting these specific cases of mixed membership

models as well as their extensions, using a general framework described by Erosheva et al. [2004] with the estimation carried out via a variational EM algorithm.

The remainder of section 1 introduces notation and formally defines the generative hierarchy assumed for discrete mixed membership models. Section 2 outlines the variational EM method used by the `mixedMem` package and discusses some of the benefits and drawbacks of the variational approach. Section 3 provides a step-by-step guide for fitting mixed membership models using `mixedMem`. For illustration, we analyze the results of the 1983 American National Election Studies Pilot Study. Finally, section 4 provides a brief conclusion.

1.2 Illustration and Notation

For a hypothetical example, consider a survey of T = 100 (the number of individuals) high school J = 3 (the number of variables) questions. Each individuals is asked to (1) rank their favorite classes (rank data), (2) select their favorite TV channel (multinomial data) and (3) indicate whether or not they are on the honor roll (Bernoulli data). Students within the same extracurricular activities/clubs might give similar responses, so in a mixed membership analysis, the latent sub-populations map to the extracurricular groupings. For concreteness, assume the following K = 4 sub-populations fully describe the students' extracurricular activities: athletics, math/science, fine arts and student government. Athletes might be more likely to rank "Anatomy" and "Health" among their favorite classes, while students involved in student government might be more likely to rank "Civics" or "History" highly. As is often the case, students may be involved with multiple clubs/activities to varying degrees. If a student is involved in student government and athletics, they might rank their favorite classes as (1. "Anatomy", 2. "History" and 3. "Health"). Thus, using a mixture model framework to classify the student solely as an athlete would fail to represent her full profile. We use λ_i to denote the distribution of membership for individual i. λ_i is a vector of K non-negative elements which sum to 1. The membership vector λ for the student above might be $(\text{athletics} = 0.7, \text{math/science} = 0.0, \text{fine arts} = 0.0, \text{student government} = 0.3)$.

We index each of the T students by $i = 1, 2, \ldots$ T; in this case T = 100. Each of the J variables are indexed by $j = 1, 2, \ldots J$ and for each variable we index the R_j number of replicates by $r = 1, 2, \ldots R_j$. For rank data, we index the $N_{i,j,r}$ ranking levels by $n = 1, 2 \ldots N_{i,j,r}$; for multinomial and binary data $N_{i,j,r} = 1$. For each individual i, variable j, replicate r, and ranking level n we denote the observed response by $X_{i,j,r,n}$. For example, if student 10's favorite classes (variable 1) are ("Anatomy", "History", "Health") and she is on the honor roll (variable 3), then $X_{i=10,j=1,r=1,n=3}$ = "Health" and $X_{i=10,j=3,r=1,n=1} = 1$.

1.3 Generative Process

More formally, for K sub-populations, we assume a mixed membership generative process as follows:
For each individual $i = 1, \ldots T$: Draw λ_i from a Dirichlet(α). λ_i is a vector of length K which indicates the degree of membership for individual i.

- For each variable $j = 1, \ldots, J$, each replicate $r = 1, \ldots, R_j$ and each ranking level $n = 1 \ldots, N_{i,j,r}$: Draw $Z_{i,j,r,n}$ from a multinomial(1, λ_i). $Z_{i,j,r,n}$ determines the sub-population which governs the response for observation $X_{i,j,r,n}$. This is sometimes referred to as the context vector because it determines the context from which the individual responds.

- For each variable $j = 1, \ldots, J$, each replicate $r = 1, \ldots, R_j$ and each ranking level $n = 1 \ldots, N_{i,j,r}$: Draw $X_{i,j,r,n}$ from the distribution parameterized by $\theta_{j,Z_{i,j,r,n}}$. Here, θ is the set of parameters which govern the observations for each sub-population. If variable j is a multinomial or rank distribution with V_j categories/candidates, $\theta_{j,k}$ is a vector of length V_j which parameterizes the responses to variable j for sub-population k. If variable j is a Bernoulli random variable, then $\theta_{j,k}$ is a value which determines the probability of success.

2 Variational EM and the `mixedMem` Package

2.1 Fitting Mixed Membership Models

When using a mixed membership model, the interest is typically in estimating the sub-population parameters θ, the Dirichlet parameter α and the latent memberships of individuals λ. Estimation of these quantities through maximum likelihood or direct posterior inference is computationally intractable in a mixed membership model, so Markov Chain Monte Carlo or variational inference techniques are used instead [Airoldi et al., 2009]. Most MCMC analyses typically require large amounts of human effort to tune and check the samplers for convergence. Furthermore, in mixed membership models, we must sample a latent membership for each individual and a context vector for each observed response; thus, a mixed membership MCMC analysis becomes very computationally expensive as the number of individuals grows. `mixedMem` circumvents these difficulties by employing a variational EM algorithm which is a deterministic and computationally attractive alternative for fitting mixed membership models. Using a variational approach allows us to fit larger models and avoids tedious human effort by approximating the true posterior and replacing the MCMC sampling procedure with an optimization problem [Beal, 2003].

2.2 Variational EM

To fit a mixed membership model, the variational EM algorithm combines variational inference (inference on an approximate posterior) to estimate the group memberships λ and a pseudo maximum likelihood procedure to estimate the group parameters θ and Dirichlet parameter α.

Instead of working with the intractable true posterior, variational inference employs a more computationally tractable approximate variational distribution. This variational distribution, denoted by Q, is parameterized by ϕ and δ as follows:

$$p(\lambda, Z|X) \approx Q(\lambda, Z|\phi, \delta) = \prod_{i}^{T} \text{Dirichlet}(\lambda_i|\phi_i) \prod_{j}^{J} \prod_{r}^{R_j} \prod_{n}^{N_{i,j,r}} \text{multinomial}(Z_{i,j,r,n}|\delta_{i,j,r,n}). \tag{1}$$

The parameters ϕ and δ can be selected to minimize the Kullback-Leibler divergence between the true posterior distribution and the variational distribution Q [Beal, 2003]. This provides an approximate distribution which can be used to carry out posterior inference on the latent variables and posterior means which can be used as point estimates for λ.

This variational distribution can also be used in an alternative pseudo-likelihood framework, to select the global parameters θ and α. Using Jensen's inequality, the variational distribution can be used to derive a function of ϕ, δ, α and θ which is a lower bound on the log-likelihood of our observations:

$$p(X|\alpha, \theta) \geq \mathbb{E}_Q \left\{ \log \left[p(X, Z, \Lambda) \right] \right\} - \mathbb{E}_Q \left[\log \left[Q(Z, \Lambda|\phi, \delta) \right] \right]. \tag{2}$$

The lower bound on the RHS of equation (2) is often called the ELBO for **E**vidence **L**ower **Bo**und. Calculating the LHS of equation (2) is intractable, but for a fixed ϕ and δ, selecting α and θ to maximize the lower bound on the RHS is a tractable alternative to maximum likelihood estimation.

It can be shown that minimizing the KL Divergence between Q and the true posterior is actually equivalent to maximizing the lower bound in equation (2) with respect to ϕ and δ [Beal, 2003]. Thus, the tasks of finding an approximate posterior distribution with respect to ϕ and δ and picking pseudo-MLE's θ and α are both achieved by maximizing the lower bound in equation (2). In practice, we maximize this lower bound through a variational EM procedure. In the E-step (variational inference), we fix θ and α and minimize the KL divergence between Q and the true posterior (this is also equivalent to maximizing the lower bound in equation (2) through iterative closed form updates to ϕ and δ. In the M-step (pseudo-MLE), we fix ϕ and δ and select the θ and α which maximize the lower bound through gradient based methods. The entire procedure iterates between the E-step and the M-step until reaching a local mode. A detailed exposition of variational inference is provided by Jaakkola [2001].

2.2.1 Label Switching in Mixed Membership Models

Mixed membership models are only identifiable up to permutations of the sub-population labels (ie simultaneously permuting the labels for all group memberships and distribution parameters). In an MCMC analysis of mixed membership models, this can require special attention if label switching is present within a sampler [Stephens, 2000]. Because variational EM is a deterministic approach, the final permutation of the labels is only dependent on the initialization points and does not require special attention during the fitting procedure. However matching group labels can still a concern for assessing model fit when comparing to some ground truth or another fitted model. `mixedMem` provides functions discussed in subsection 3.1 for dealing with the label permutations to facilitate post-processing.

2.2.2 Variational EM Caveats

The computational benefits of variational EM, however, do not come for free. The lower bound in equation (2), is generally not a strictly convex function, so only convergence to a local mode, not the global mode, is guaranteed [Wainwright and Jordan, 2008]. If prior knowledge exists about a specific problem, initializing θ and α near plausible values is helpful in ensuring that the EM algorithm reaches a reasonable mode. In general though, starting from multiple initialization points and selecting the mode with the largest ELBO is highly recommended. This is also an area where the post-processing tools discussed in subsection 3.1 can be useful for determining if candidate modes found from different initialization points are conceptually different.

Variational inference also lacks any guarantees on the quality of our approximation. Erosheva et al. [2007] show that a mixed membership analysis of survey data using a MCMC and variational approach agree well, and, in practice, we see that variational inference provides reasonable and interpretable results in mixed membership models [Blei et al., 2003, Erosheva et al., 2004, Airoldi et al., 2009]. However, there are no theoretical guarantees on how good or bad our approximation ultimately is.

3 Example: Fitting Political Survey Data with `mixedMem`

For demonstration, we present a `mixedMem` analysis of political opinion survey data previously analyzed by Gross and Manrique-Vallier [2014] as well as Feldman [1988]. Within this context, identified latent sub-populations might map to ideological blocs. Since individuals often hold to political ideologies to varying degrees, a mixed membership model is particularly appropriate. We utilize the mixed membership model as specified by Gross and Manrique-Vallier and discussed more generally in section 1.3. The model assumes 3 latent sub-populations (K = 3) with 19 observed multinomial variables with 1 replicate each. Gross and Manrique-Vallier specify a fully Bayesian approach, placing prior distributions on both θ and α and utilize

MCMC to estimate the model. This allows for posterior inference on both the latent memberships as well as on θ and α. Our analysis using `mixedMem` will fit the model using the variational EM algorithm which allows for posterior inference on λ and Z, but only yields point estimates for α and θ. Nonetheless, we will show that the two methods yield comparable results and very similar interpretations.

In the 1983 American National Election Survey Pilot [ANES, 1999], each individual was prompted with various opinion-based statements and was asked to report their agreement with the statement using the categories: "strongly agree", "agree but not strongly", "can't decide", "disagree but not strongly", and "strongly disagree". For example, one statement was "Any person who is willing to work hard has a good chance of succeeding". We specifically study 19 of the statements which were selected by Feldman [1988] and reanalyzed by [Gross and Manrique-Vallier, 2014]. The 19 statements can be grouped into 3 overarching themes: Equality (abbreviated "EQ" in the data set variable names), Economic Individualism (abbreviated "IND") and Free Enterprise (abbreviated "ENT") [Feldman, 1988]. Following the original analysis of Gross and Manrique-Vallier [2014], we include the 279 complete responses and combine categories "agree" with "strongly agree" and "disagree" with "strongly disagree", leaving the 3 possible responses "agree" = 0, "can't decide" = 1, and "disagree" = 2 to avoid overparameterization. The data is included in `mixedMem` as `ANES`; the full text statement for each variable can be accessed through `help(ANES)`.

A brief exploratory analysis, shows that of the $279 \times 19 = 5301$ total responses, 3295 responses are "agree" 1907 are "disagree" and only 99 are "can't decide".

```
library(mixedMem)

## Loading required package:  gtools

data(ANES)
# Dimensions of the data set: 279 individuals with 19 responses each
dim(ANES)

## [1] 279  19

# The 19 variables and their categories
# The specific statements for each variable can be found using help(ANES)
# Variables titled EQ are about Equality
# Variables titled IND are about Econonic Individualism
# Variables titled ENT are about Free Enterprise
colnames(ANES)

##  [1] "EQ1"  "EQ2"  "EQ3"  "EQ4"  "EQ5"  "EQ6"  "EQ7"  "IND1" "IND2" "IND3"
## [11] "IND4" "IND5" "IND6" "ENT1" "ENT2" "ENT3" "ENT4" "ENT5" "ENT6"

# Distribution of responses
table(unlist(ANES))

##
##    0    1    2
## 3295   99 1907
```

Step 1: Initializing the `mixedMemModel` Object

To fit a mixed membership model, we must first initialize a `mixedMemModel` object using the class constructor. The `mixedMemModel` object contains the dimensions of our mixed membership model, the observed data and initialization points for α and θ. Creating a `mixedMemModel` object provides a vehicle for passing this information to the `mmVarFit` function in step 2. This is similar to how one might specify the formula for an `lm` object. Although initialization points for ϕ and δ can be passed to the constructor as well, these by default are initialized uniformly across the sub-populations, and unless there is very strong prior knowledge, we recommend that the default values be used. For this particular model, all the variables are multinomials; an example showing an initialization of mixed data types can be accessed through `help(mixedMemModel)`.

As mentioned previously, because the lower bound function is not convex, different initializations may result in convergence to different local modes. After initializing at various point, we found that the initialization of $\alpha = (.2, .2, .2)$ and $\theta \sim$ Dirichlet(.8) using seed 123 resulted in the highest lower bound at convergence.

```
# Sample Size
Total <- 279
# Number of variables
J <- 19
# we only have one replicate for each of the variables
Rj <- rep(1, J)
# Nijr indicates the number of ranking levels for each variable.
# Since all our data is multinomial it should be an array of all 1s
Nijr <- array(1, dim = c(Total, J, max(Rj)))
# Number of sub-populations
K <- 3
# There are 3 choices for each of the variables ranging from 0 to 2.
Vj <- rep(3, J)
# we initialize alpha to .2
alpha <- rep(.2, K)
# All variables are multinomial
dist <- rep("multinomial", J)
# obs are the observed responses. it is a 4-d array indexed by i,j,r,n
# note that obs ranges from 0 to 2 for each response
obs <- array(0, dim = c(Total, J, max(Rj), max(Nijr)))
obs[,,1,1] <- as.matrix(ANES)

# Initialize theta randomly with Dirichlet distributions
set.seed(123)
theta <- array(0, dim= c(J,K,max(Vj)))
for(j in 1:J)
{
    theta[j,,] <- rdirichlet(K,rep(.8,Vj[j]))
}

# Create the mixedMemModel
# This object encodes the initialization points for the variational EM algorithim
```

```
# and also encodes the observed parameters and responses
initial <- mixedMemModel(Total = Total, J = J, Rj = Rj,
                         Nijr = Nijr, K = K, Vj = Vj, alpha = alpha,
                         theta = theta, dist = dist, obs = obs)
```

Step 2: Fitting the Model

Now we can fit the model using the `mmVarFit` function. When we call `mmVarFit`, the function first checks for internal consistency in the input model, `initial` (i.e., it checks whether the dimensions of the observation match the specified number of variables, etc). If the model check passes, the function begins to iterate through the E-step and M-step described in subsection 2.2. When the algorithm converges[1], `mmVarFit` prints the value of the lower bound at convergence as well as the number of EM iterations needed for convergence.

```
computeELBO(initial)

## [1] -16810.18

st = proc.time()
out <- mmVarFit(initial)

## [1] "Model Check: Ok!"
## [1] "<== Beginning Variational Inference ==>"
## Fit Complete! Elbo: -3122.89 Iter: 103

end = proc.time()
computeELBO(out)

## [1] -3122.886

time = end - st
```

Notice that the lower bound (our objective function) is -1.681018×10^4 at the initialized points and is -3122.89 at convergence. On a laptop (quad-core 2.4 GHZ with 12 GB of RAM), fitting the entire model took only 10.69 seconds, significantly faster than a full MCMC analysis. We can also view a quick summary of the fit model using the `summary` function.

```
summary(out)

## ==Summary for Mixed Membership Model==
## Total: 279  K: 3  ELBO: -3122.89
##
##  Variable   Variable Type     Replicates     Categories
##    1        multinomial       1              3
##    2        multinomial       1              3
##    3        multinomial       1              3
```

[1]The maximum iterations before the algorithim terminates is 500

```
##    4        multinomial       1              3
##    5        multinomial       1              3
##    6        multinomial       1              3
##    7        multinomial       1              3
##    8        multinomial       1              3
##    9        multinomial       1              3
##    10       multinomial       1              3
##    11       multinomial       1              3
##    12       multinomial       1              3
##    13       multinomial       1              3
##    14       multinomial       1              3
##    15       multinomial       1              3
##    16       multinomial       1              3
##    17       multinomial       1              3
##    18       multinomial       1              3
##    19       multinomial       1              3
```

3.1 Step 3: Post-Processing

In many cases, we may be interested in comparing our fitted model to some ground truth or comparing various runs with different starting values. Since the model is invariant to label permutations, we use the `findLabels` function to match our sub-population labels to the ground truth labels to facilitate comparison. In this case, we are interested in comparing our fitted values with the posterior means from the original MCMC analysis of Gross and Manrique-Vallier [2014]. The `findLabels` function finds the permutation of labels which minimizes the weighted squared error shown in equation (3). Once we have found the optimal permutation, we then permute the sub-population labels using the `permuteLabels` function.

$$Loss = \sum_{i}^{T} \sum_{j}^{J} \frac{\hat{\alpha}_k}{\hat{\alpha}_0} \sum_{v}^{V_j} (\hat{\theta}_{j,k,v} - \theta_{j,k,v})^2 \tag{3}$$

Where $\hat{\alpha}_0 = \sum_k \alpha_k$.

```
#read in GM estimates
data(gmv_theta)
#
optimal.perm <- findLabels(out, gmv_theta)
# display the permutation as well as the weighted squared error loss
optimal.perm

## $perm
## [1] 3 2 1
##
## $loss
## [1] 5.205028

# save object with permuted labels
out.permute <- permuteLabels(out, optimal.perm$perm)
```

3.2 Step 4: Interpretation

Visualizations of the fitted model can be highly context-specific and each of the estimated quantities can be extracted from the output model for further analysis tailored to the scientific question of interest.

```
# The estimated quantities can be exctracted from the output model
names(out)

##  [1] "Total" "J"     "Rj"    "Nijr"  "K"     "Vj"    "alpha" "theta"
##  [9] "phi"   "delta" "dist"  "obs"

out$alpha

## [1] 0.02921846 3.44769371 3.61722884
```

In the following subsections, we show a few visualizations and tools which may be of general interest.

3.2.1 Visual Representation of $\hat{\theta}$

We can now plot our estimated $\hat{\theta}$ values and compare them to the values found in the original analysis. In figures 1 and 2, each row of plots represents an individual question, the columns represent each of the identified sub-populations and the plots show the sub-population's response probability for each variable. The black bars indicate the values estimated using `mixedMem` and the green dots indicate the values reported by [Gross and Manrique-Vallier, 2014]. Note that Gross and Manrique-Vallier do not report values for group 3 for reasons discussed in subsection 3.2.2. Recall that 0 indicates "agree" and 2 indicates "disagree".

```
vizTheta(out.permute, gmv_theta, varNames = colnames(ANES),
         groupNames = c("Conservative", "Liberal", "Undecided"))
```

When examining the estimates of $\hat{\theta}$, we can see from figure 3 that of the 19 variables, the statements most likely to elicit agreement from sub-population 1 are statement `IND1` ("Any person who is willing to work hard has a good chance of succeeding"), `IND3` ("Most people who don't get ahead should not blame the system; they really have only themselves to blame") and `ENT1` ("The less government gets involved with business and the economy, the better off this country will be"). Using traditional political ideology labels, sub-population 1 could be identified as the conservative bloc. In figure 3, bars shown in green indicate question to which sub-population 1 is more likely to agree ($\theta_{j,1,1} \geq .5$) and bars shown in red indicate questions to which the sub-population is not likely to agree.

```
pop1VarOrder = colnames(ANES)[order(out.permute$theta[,1,1], decreasing = T)]
pop1VarAgree = sort(out.permute$theta[,1,1], decreasing = T)
barplot(height = pop1VarAgree, names.arg = pop1VarOrder,
        main = "Propensity to Agree",
        cex.names = .7, las = 2, xlab = "Value Statements",
```

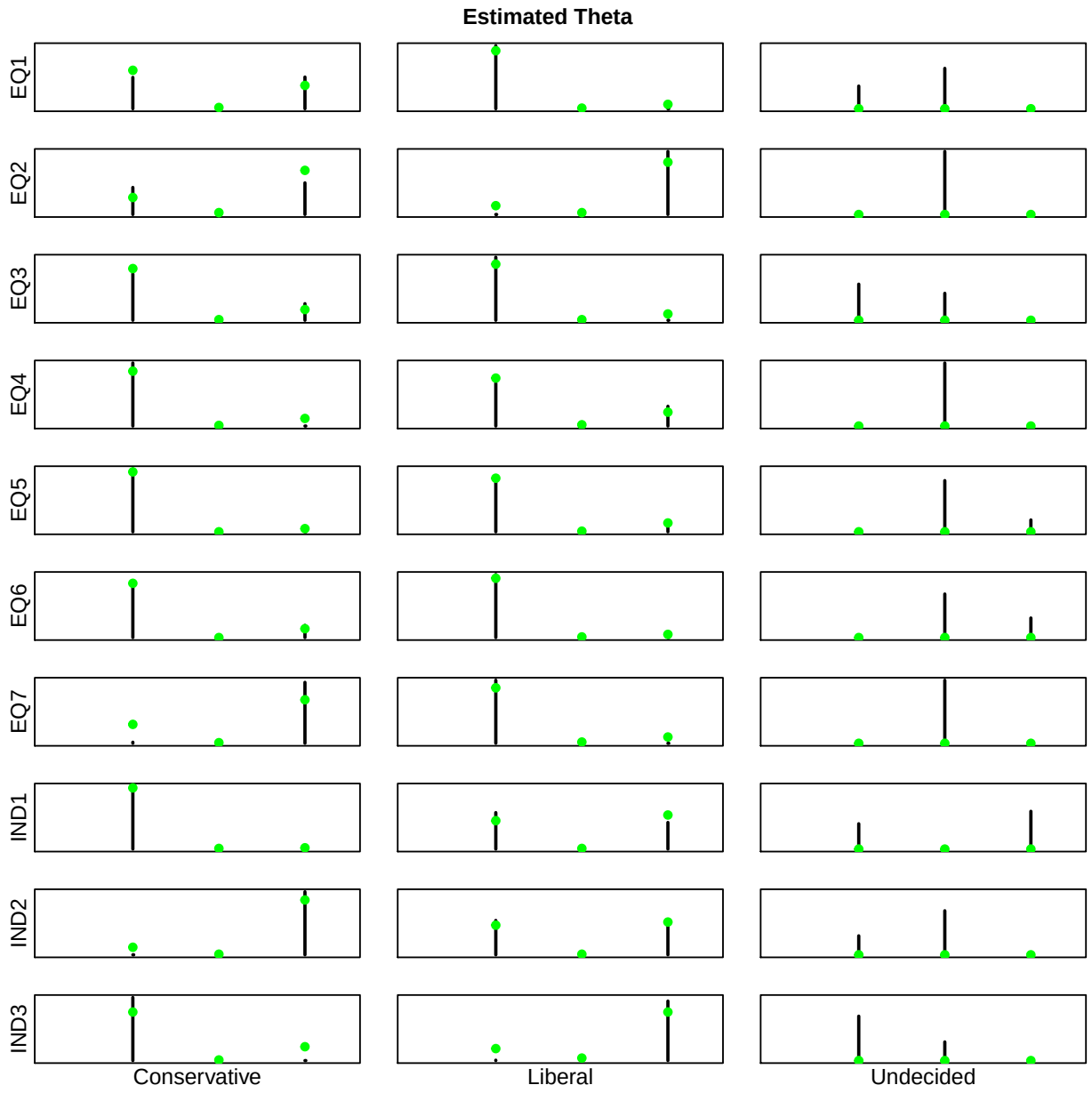

Figure 1: Individuals were asked to indicate their level of agreement with 19 opinion based statements. The fitted multinomial response probabilities for each ideology bloc to each of the 19 statements are displayed. On the horizontal axis, 0 indicates agree, 1 indicates can't decide, and 2 indicates disagree. The estimates from our variational analysis are denoted by the black bars, the Gross and Manrique-Vallier results using MCMC are shown by the green circles

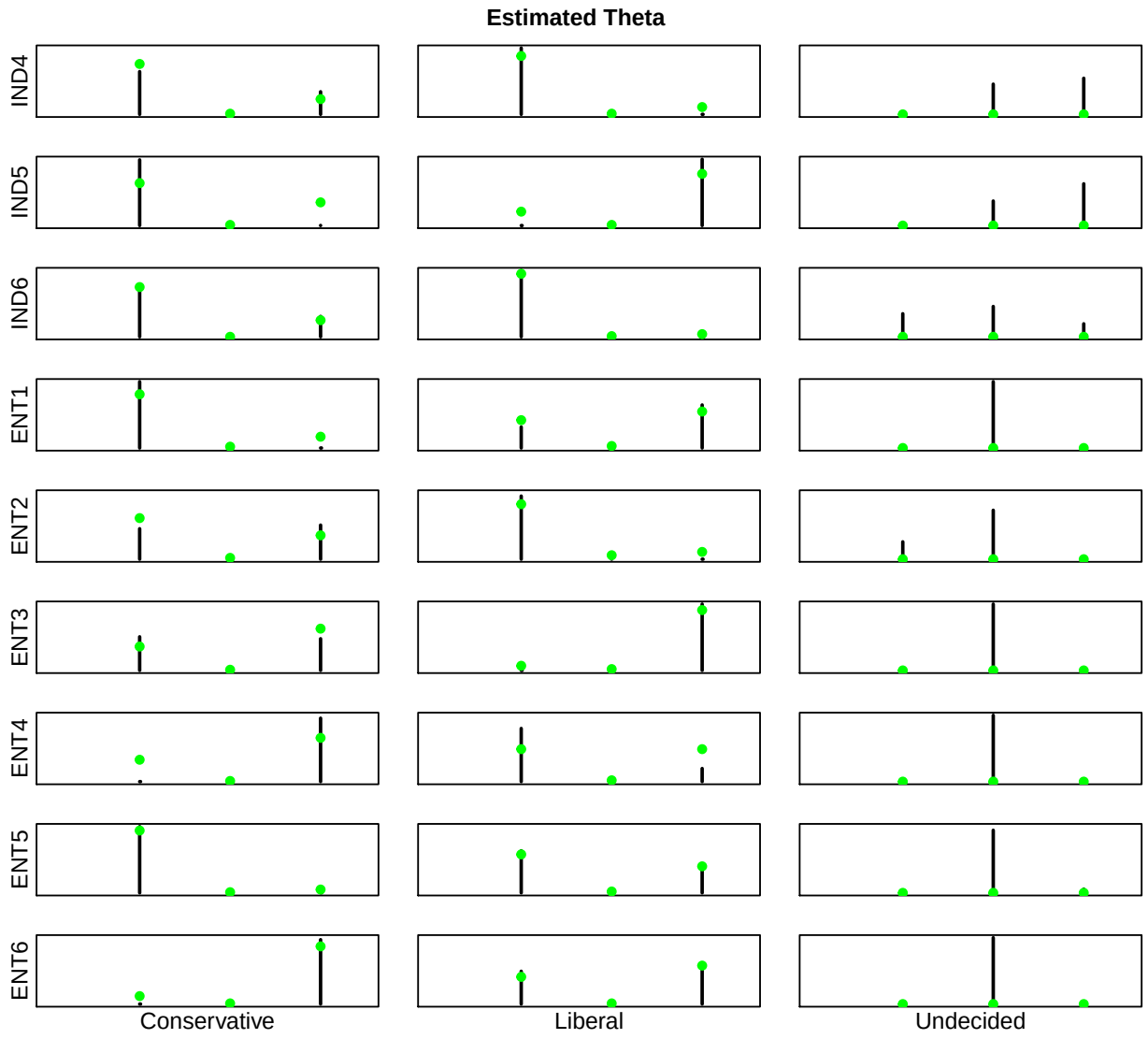

Figure 2: Individuals were asked to indicate their level of agreement with 19 opinion based statements. The fitted multinomial response probabilities for each ideology bloc to each of the 19 statements are displayed. On the horizontal axis, 0 indicates agree, 1 indicates can't decide, and 2 indicates disagree. The estimates from our variational analysis are denoted by the black bars, the Gross and Manrique-Vallier results using MCMC are shown by the green circles

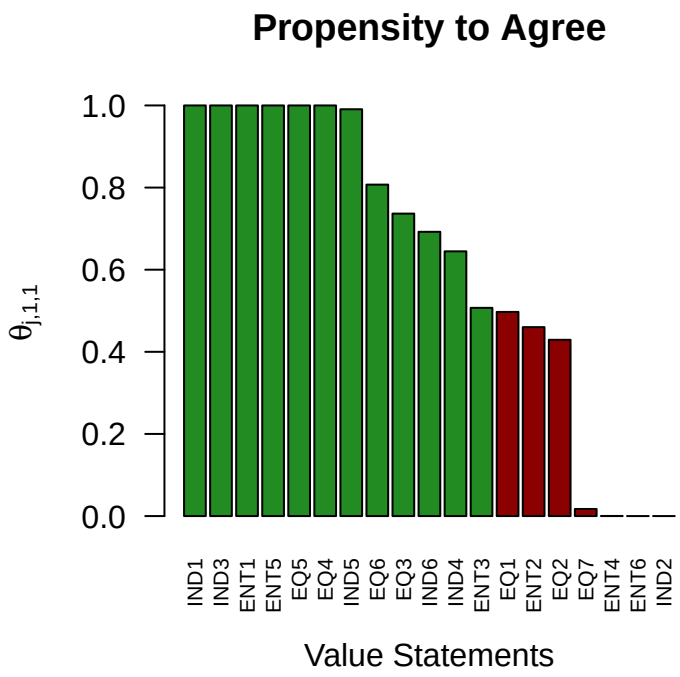

Figure 3: Propensity to agree with each opinion-based statement for the conservative bloc

```
        ylab = expression(paste(theta["j,1,1"])),
        col =ifelse(pop1VarAgree > .5, "forestgreen", "darkred"))
```

We can see from figure 4 that for sub-population 2, the statements most likely to elicit agreement are statements `EQ7` ("One of the big problems in this country is that we don't give everyone an equal chance"), `IND6` ("Even if people try hard, they often cannot reach their goals ") and `EQ1` ("If people were treated more equally in this country, we would have many fewer problems "). Sub-population 2 could be identified as the liberal bloc.

```
pop2VarOrder = colnames(ANES)[order(out.permute$theta[,2,1], decreasing = T)]
pop2VarAgree = sort(out.permute$theta[,2,1], decreasing = T)
barplot(height = pop2VarAgree,
        names.arg = pop2VarOrder, main = "Propensity to Agree",
        cex.names = .7, las = 2, xlab = "Value Statements",
        ylab = expression(paste(theta["j,2,1"])),
        col =ifelse(pop2VarAgree > .5, "forestgreen", "darkred"))
```

We also observe, that sub-population 3 has a much higher propensity to respond "can't decide" then sub-populations 1 or 2. The 5 individuals with particularly large membership in group 3 responded "can't decide" 42 times. This is particularly salient since there are only 99 "can't decide" responses in the entire sample. Thus, sub-population 3 could be identified as the undecided bloc.

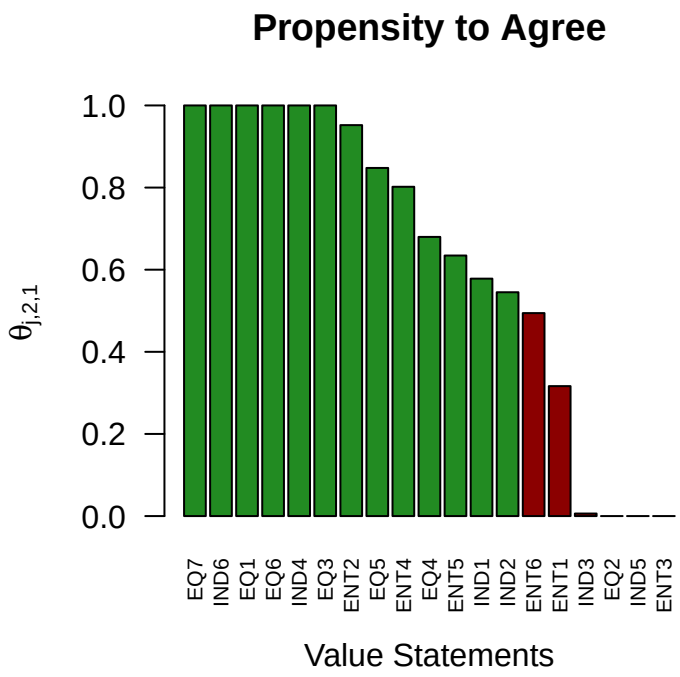

Figure 4: Propensity to agree with each opinion-based statement for the liberal bloc

```
# Point estimates for lambda
lambda.point <- out.permute$phi/rowSums(out.permute$phi)
# number of individuals which exhibit more than .3 degree of membership
# in the undecided group
sum(lambda.point[,3]>=.3)

## [1] 5

# number of can not decide responses from those with high membership in undecided group
sum(ANES[which(lambda.point[,3]>=.3),]==1)

## [1] 42
```

3.2.2 Interpretation of $\hat{\alpha}$

When examining the fitted value of $\hat{\alpha}$, we see that the conservatives and liberals are the 2 dominant groups with the undecided group comprising a much smaller portion of the population. Also, the large values of $\hat{\alpha}_1$ and $\hat{\alpha}_2$, indicate a high level of mixing between those sub-populations. Since the estimated relative frequency of the undecided bloc is less than 1%, we focus our interpretation on the other two blocs dominant blocs.

```
relativeFrequency = out.permute$alpha/sum(out.permute$alpha)
```

	Conservatives	Liberals	Undecided
Estimated Alpha	3.617	3.448	0.029
Estimate Relative Frequency	0.510	0.486	0.004

Table 1: Variational Estimates of Alpha

3.2.3 Using Hellinger Distance to Determine Defining Characteristics for Each Bloc

We might also be interested in seeing which questions are the most polarizing (the questions to which a conservative is most likely to respond differently than a liberal). These particular value statements provide insight into defining what makes a conservative and what makes a liberal. We use the Hellinger distance, a measure of the difference between distributions, to compare the conservative and liberal response probabilities for each question. Hellinger distance is defined as

$$H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_i (\sqrt{p_i} - \sqrt{q_i})^2}. \tag{4}$$

A Hellinger distance close to 1 indicates that two probability distributions are dissimilar, while a Hellinger distance close to 0 indicates that the two probaility distirbutions are similar. We can see in figure 5 that the three most polarizing questions, as measured by the Hellinger distance, are statements `IND5` ("If people work hard, they almost always get what they want"), `IND3` ("Most people who don't get ahead should not blame the system; they really have only themselves to blame") and `EQ7` ("One of the big problems in this country is that we don't give everyone an equal chance"). Thus, it would seem that the most polarizing issues in 1983 revolved around access and opportunity for advancement.

```
hellingerDist = (1/sqrt(2))*sqrt(rowSums((sqrt(out.permute$theta[,1,])
                                          - sqrt(out.permute$theta[,2,]))^2))
barplot(sort(hellingerDist, decreasing = T),
        names.arg = colnames(ANES)[order(hellingerDist, decreasing = T)],
        main = "Hellinger Distance",
        cex.names = .7, las = 2, ylab = "Hellinger Distance",
        ylim = c(0,1))
mtext("Between Conservatives and Liberals")
```

```
colnames(ANES)[order(hellingerDist, decreasing = T)][1:3]

## [1] "IND5" "IND3" "EQ7"
```

3.3 Visualizing Group Dispersion

For the groups memberships for individuals, we use the posterior mean, $\frac{\hat{\phi}_i}{\sum_k \hat{\phi}_{i,k}}$ as point estimates for λ_i. Plotting the poster mean of membership in the conservative bloc below, we can see a fair amount of intra-individual mixing. 139 out of the 279 individuals have estimated memberships of at least 40% in both the

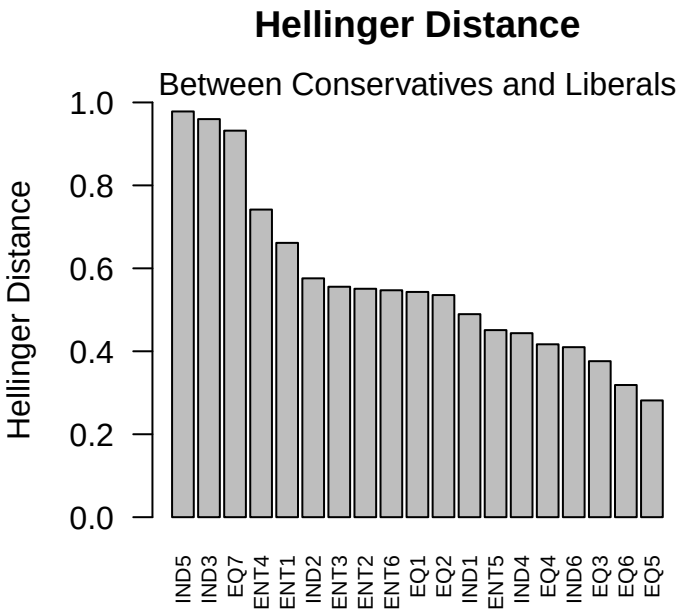

Figure 5: Using Hellinger distance indicates that the most polarizing issues involved opportunity for advancement

conservative and liberal blocs. This is not particularly uprising since we observed relatively large values of $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

```
estimatedLambda = out.permute$phi/rowSums(out.permute$phi)
# number of individuals with at least 40% membership in
# both conservative and liberal blocs
sum(estimatedLambda[,1]>.4 & estimatedLambda[,2] >.4)

## [1] 139
```

We can also plot the "empirical CDF" of estimated membership in conservative bloc as shown in figure 6. Since the marginal distribution of a Dirichlet distribution is a beta, we can also plot the 95% credible intervals for the posterior membership in the conservative bloc. We observe that there is still a relatively large amount of uncertainty in the estimated membership of each individual.

```
index = order(estimatedLambda[,1])

# variance of posterior membership in conservative bloc
var.Mem = out.permute$phi[,1]*(rowSums(out.permute$phi)-
                                 out.permute$phi[,1])/
  (rowSums(out.permute$phi)^2*(rowSums(out.permute$phi)+1))
# plot posterior means
```

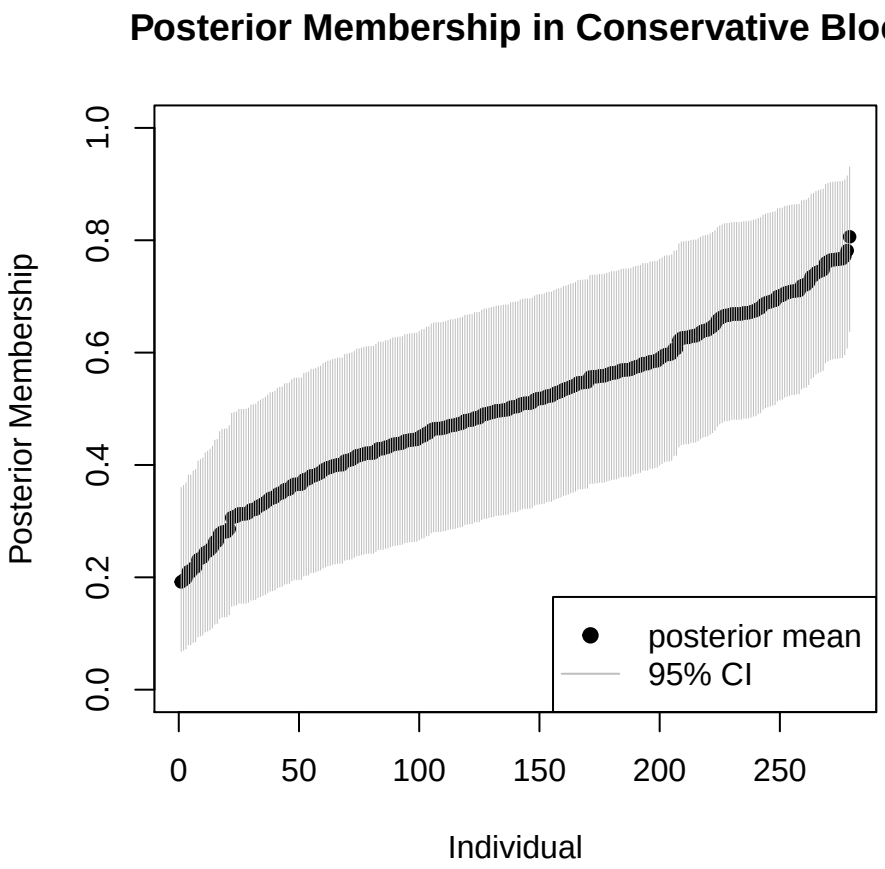

Figure 6: We observe a relatively high rate of intra-individual mixing

```
plot(sort(estimatedLambda[,1]), pch = 19,
     main = "Posterior Membership in Conservative Bloc",
     ylab = "Posterior Membership", xlab = "Individual",
     cex = .8, ylim = c(0,1))

# marginal distirbution of Dirichlet, is Beta distribution, so we can get posterior CI
# plot posterior 90% CI
ci_up = qbeta(.975, out.permute$phi[index,1], rowSums(out.permute$phi[index,c(2:3)]))
ci_low = qbeta(.025, out.permute$phi[index,1], rowSums(out.permute$phi[index,c(2:3)]))
segments(x0 = c(1:out$Total), y0 = ci_up, y1 = ci_low, col = "gray", lwd = .3, lty = 1)
legend("bottomright", legend = c("posterior mean", "95% CI"), pch = c(19, NA),
       lty = c(NA,1), col = c("black", "gray"))
```

3.4 Comparison of MCMC and Variational Results

Comparing the results of the MCMC analysis by Gross and Manrique-Vallier [2014] to the results from our variational analysis, we see they are very similar. In both analyses, we identify two dominant profiles- a

conservative bloc and a liberal bloc- as well as a much smaller undecided faction. As can be seen in figure 2, the variational estimates of θ agree well with the MCMC estimates.

Since Gross and Manrique-Vallier utilize a fully Bayesian specification, they are able to utilize a hypothesis testing framework with θ to find the most polarizing issues. Because the variational analysis only provides point estimates of θ, we used Hellinger distances instead. However, both analyses agree that the three most polarizing statements are `IND5`, `IND3` and `EQ7`.

Although the broad interpretation and estimates of θ agree, we do see differences in the estimates of α. Gross and Manrique-Vallier report a posterior mean of $\alpha = (0.462, 0.285, 0.018)$ yielding relative frequencies of $(60.4\%, 37.3\%, 2.3\%)$ for the conservative, liberal and undecided bloc respectively. Although MCMC ordering of the relative frequency matches the ordering of the variational estimates, this implies a much lower level of intra-individual mixing and as well as a higher relative frequency for the conservative bloc.

4 Conclusion

In this tutorial, we only briefly introduced the ideas of mixed membership models. For more interested readers, Airoldi et al. [2014] provide a much deeper exposition of mixed membership models as well as a variety of different applications.

We also provide a step-by-step guide for using `mixedMem` as well as some sample visualizations/interpretations which may be helpful to the user. From the political survey example, we see that the use of variational inference largely agrees with the more complicated MCMC procedure, and still provides reasonable and interpretable results.

By providing an R package for fitting mixed membership models, we aim to aid researchers who are studying problems where a mixed membership analysis is compelling, but have been otherwise dissuaded by the computational difficulties. We hope that this package extends the use of mixed membership models to a variety of new disciplines and interesting scientific problems.

References

Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. In *Advances in Neural Information Processing Systems*, pages 33–40, 2009.

Edoardo M. Airoldi, David Blei, Elena A. Erosheva, and Stephen E. Fienberg. *Handbook of Mixed Membership Models and Their Applications*. Chapman and Hall, 2014.

American National Election Studies ANES. National election studies, 1983 pilot election study, 1999. URL `http://www.electionstudies.org/studypages/1983pilot/1983pilot.htm`.

Matthew James Beal. *Variational algorithms for approximate Bayesian inference*. PhD thesis, University of London, 2003.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

Elena Erosheva, Stephen Fienberg, and John Lafferty. Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5220–5227, 2004.

Elena A Erosheva, Stephen E Fienberg, and Cyrille Joutard. Describing disability through individual-level mixture models for multivariate binary data. *The Annals of Applied Statistics*, 1(2):346, 2007.

Stanley Feldman. Structure and consistency in public opinion: The role of core beliefs and values. *American Journal of Political Science*, pages 416–440, 1988.

Isobel Claire Gormley and Thomas Brendan Murphy. A grade of membership model for rank data. *Bayesian Analysis*, 4(2):265–295, 2009.

Justin Gross and Daniel Manrique-Vallier. *Handbook of Mixed Membership Models and Its Applications*, chapter A Mixed Membership Approach to the Assessment of Political Ideology from Survey Responses, pages 119–139. Chapman & Hall/CRC Press, 2014.

Tommi S Jaakkola. 10 tutorial on variational approximation methods. *Advanced mean field methods: theory and practice*, page 129, 2001.

Jonathan K Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.

Matthew Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809, 2000.

Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.