2

A Tale of Two (Types of) Memberships: Comparing Mixed and Partial Membership with a Continuous Data Example

Jonathan Gruhl

Department of Statistics, University of Washington, Seattle, WA 98195, USA

Elena A. Erosheva

Department of Statistics, University of Washington, Seattle, WA 98195, USA

CONTENTS

Mixed membership models such as the Grade of Membership and latent Dirichlet allocation models have primarily focused on the analysis of binary and categorical data. In this chapter, we will focus on exploring the performance of two different types of membership models with continuous data: one that has a classic mixed membership structure and one that has a partial membership structure. The Bayesian partial membership model was recently proposed by Heller et al. (2008) as a promising alternative to mixed membership motivated by continuous data. The Bayesian partial membership model based on exponential family distributions allows for computationally efficient modeling of a variety of data types. Heller et al. (2008) demonstrated a partial membership analysis of a discrete dataset. In this work, we use a dataset that has a collection of continuous variables describing NBA (National Basketball Association) players and their playing styles as a motivating example. Although NBA players are typically assigned to one of five player positions, the language used to describe players and playing styles is often suggestive of individual-level mixtures. In this chapter, we compare the exponential family form of the Bayesian partial membership model with the general mixed membership model on simulated binary and continuous data. We then extend the partial membership framework to account for correlated membership scores. Based on the

properties of the two types of models and the nature of the NBA data, we argue for choosing a partial membership model over a mixed membership model in this case. We show how the NBA players can be modeled as individual-level mixtures using the correlated partial membership model. To our knowledge, this is the first individual-level mixture analysis of continuous data.

2.1 Introduction

Mixture models provide a model-based approach to clustering. Population-level mixture models describe a population as a collection of subpopulations where each individual (or observational unit) belongs exclusively to one of the subpopulations (Lazarsfeld and Neil, 1968). Individual-level mixture models, on the other hand, allow each individual to belong to multiple subpopulations at once, with varying degrees of membership among individuals (Woodbury et al., 1978; Pritchard et al., 2000; Blei et al., 2003; Erosheva, 2002). Because the instance of individuals belonging exclusively to one subpopulation is a special case of individuals belonging simultaneously to multiple subpopulations, individual-level mixture models can be viewed as a relaxation of population-level mixtures such as finite mixture or latent class models.

The family of mixed membership models constitutes the predominant means of employing individual-level mixture models. At a high level, the mixed membership model assumes that data arise from individual-specific distributions that are arithmetic averages of the subpopulation distributions with individual-specific weights. Heller et al. (2008) formulated an alternative structure for individual-level mixtures, the Bayesian partial membership model, where the data can be viewed as arising from a (normalized) weighted geometric average of the subpopulation distributions with individual-specific weights.

When the subpopulation distributions are of exponential family form, the partial membership model allows for computationally efficient, individual-level mixture modeling of a variety of data types. In this chapter, we concentrate on the exponential family form of the partial membership model and compare this model to corresponding mixed membership models for the binary data case and the continuous data case. We highlight the differences in the data-generating behavior between the two types of models which have connection to the work of Galyardt (2014).

To demonstrate an individual-level mixture model analysis with continuous data, we use (NBA) National Basketball Association player statistics from the 2010–11 season (Hoopdata, 2012). The case of continuous data is of particular interest as existing individual-level mixture models have given less attention to continuous data. Even though the general class of mixed membership models (Erosheva, 2002) can accommodate any type of outcome (discrete or continuous), or even a mix of different types of outcomes in a model, the early independent developments have been motivated by discrete data problems whether in genetics, medicine, or computer science. Likewise, existing applications of mixed membership models primarily focus on binary, multinomial, and rank data: medical classification based on observed symptoms (Woodbury et al., 1978), counts of words in documents (Blei et al., 2003), responses to binary or multiple choice survey items such as disability manifestations (Erosheva et al., 2007), voter rankings of political candidates (Gormley and Murphy, 2009), counts of features present in an image (Wang et al., 2009), presence or absence of interactions between units (Airoldi et al., 2008), etc. A mixed membership analysis of continuous gene expression data with the latent process decomposition model by Rogers et al. (2005) is an exception. One reason for the continued focus of mixed membership models on discrete data is that little is known about this type of modeling for continuous data, and that basic examples of mixed membership for continuous data do not seem realistic.

The rest of the chapter is organized as follows. We provide more background on the NBA player data in Section 2.2. Section 2.3 provides a review of the mixed membership model and the partial

membership model (no implied connection to the partial membership model in Erosheva, 2004). We compare the models and their applications to binary and continuous data in Section 2.4. Our comparison of these models for the continuous data case helps explicate our decision to use the partial membership for the NBA data analysis. In Section 2.5, we introduce an extension of the partial membership model that allows us to accommodate correlations among class membership scores. In Section 2.6, we analyze NBA playing styles using the correlated partial membership model.

2.2 Compositional Playing Styles of NBA Players

In the New York Times basketball blog *Off The Dribble*, Joshua Brustein highlighted NBA-related research presented at the 2012 MIT Sloan Sports Analytics Conference (Brustein, 2012). Team chemistry and construction were recurring themes in the research with the intent of understanding how team chemistry and construction might relate to winning. In understanding the team construction process, comparing it across teams, and ultimately relating it to game outcomes, it is helpful to be able to group players by playing style and/or ability.

Typically, basketball players are assigned to one of five positions: point guard (PG), shooting guard (SG), small forward (SF), power forward (PF), and center (C). Some players may play multiple positions. For instance, some players may play both the point guard and shooting guard positions or both the small forward and power forward positions. NBA observers may commonly use a more informal typology of players with three categories that consolidate the above positions by physical attributes and function on the court: point guard, wings (shooting guards and small forwards), and bigs (power forwards and centers). However, current positions and player assignments to those positions may not fully reflect the variety of playing styles (Lutz, 2012). To classify players based on their playing style as reflected in their statistics, Lutz (2012) carried out a model-based cluster analysis of players based on their season statistics. We would like to take a different approach and, rather than assign players to strictly one playing style or identify clusters that are themselves mixtures of more pure clusters, assume that players themselves demonstrate compositions of different pure playing styles. This assumption is intuitively plausible. For instance, the term "combo guard" is regularly used to describe a player who combines the skills and the playing style of a typical point guard and a typical shooting guard. As a result, we would like to use an individual-level mixture model for our analysis of the NBA data.

To characterize players, we consider 13 different statistics from the 2010–11 NBA season available on hoopdata.com (Hoopdata, 2012). Our dataset is composed of 332 players who had played 30 or more games and averaged 10 or more minutes per game. We selected 13 statistics that characterize different elements of players' styles; these largely overlap with the statistics used by Lutz (2012) in a model-based cluster analysis of similar data.

The variables in our dataset include: minutes played per game, percent of made field goals that are assisted, assist rate, turnover rate, offensive rebound rate, defensive rebound rate, steals per 40 minutes, blocks per 40 minutes, and number of shots attempted per 40 minutes at each of the following locations: at the rim, from 3–9 feet, from 10–15 feet, from 16–23 feet, and beyond the 3-point line. All of the variables are continuous, but some, such as minutes played per game (maximum of 48) or percent of field goals made (0–100), are restricted in their range.

In addition to these variables, Lutz (2012) also included the number of games played as another statistic in the cluster analysis. We elected not to use this variable as it is likely to be influenced by events such as injuries that may have little connection to a player's style. Table 2.1 lists the variables, their abbreviations, and formulas of calculated statistics in our dataset.

Figures 2.1 and 2.2 display two bivariate scatterplots for selected player statistics. The data pat-

TABLE 2.1
Variables, abbreviations, and formulas (if calculated).

Variable	Description and Formula
Min	Minutes played per game
% Ast	Percent of made field goals that are assisted $\frac{\text{field goals that are assisted}}{\text{total made field goals}}$
AR	Assist Ratio $\frac{\text{Assists} \times 100}{\text{FGA} + (\text{FTA} \times .44) + \text{Turnovers}}$
TOR	Turnover Ratio $\frac{\text{Turnovers} \times 100}{\text{FGA} + (\text{FTA} \times .44) + \text{Turnovers}}$
ORR	Offensive Rebound Rate $\frac{100 \times (\text{Player ORebs} \times (\text{Team Min}/5))}{(\text{Player Min} \times (\text{Team ORebs} + \text{Opp DRebs}))}$
DRR	Defensive Rebound Rate $\frac{100 \times (\text{Player DRebs} \times (\text{Team Min}/5))}{(\text{Player Min} \times (\text{Team DRebs} + \text{Opp ORebs}))}$
Rim	Attempted field goals at the rim per 40 minutes
Close	Attempted field goals from 3-9 feet per 40 minutes
Medium	Attempted field goals from 10-15 feet per 40 minutes
Long	Attempted field goals from 16-23 feet per 40 minutes
3s	3-point field goals attempted per 40 minutes
Stls	Steals per 40 minutes
Blks	Blocks per 40 minutes

terns presented in Figures 2.1 and 2.2 are typical of other bivariate scatterplots in this dataset (not shown). The shapes of the plotted points indicate the player's position. In the list of positions in the legend, the positions 'G', 'GF,' and 'F' are listed in addition to the five main positions listed earlier. Hoopdata.com uses these designations in their positional assignments to describe players who regularly play multiple positions. G (guard) is typically used to describe a player who plays both point guard and shooting guard, GF (guard-forward) to describe a player who plays both shooting guard and small forward, and F (forward) to describe a player who plays both small forward and power forward.

Figure 2.1 plots the assist and turnover ratios of the players. The data appear to fan out from the lower left corner, adopting an almost triangular shape. Within this shape, we can see some patterns. Players designated as point guards and guards dominate the points in the upper right. Players manning the forward, power forward, and center positions generally appear to have low assist ratios and span the range of turnover ratios, comprising the points lining the left side of the plot.

Figure 2.2 presents the corresponding plot for defensive rebound rate and 3-point field goals attempted per 40 minutes. We see a different pattern in this data with a clear cluster of players comprised of forwards, power forwards, and centers that rarely attempt 3-point field goals. Separate from this cluster is a cloudlike structure of points that tends to shoot some 3-pointers. Within the cloud, we see that point guards tend to have lower defensive rebound rates while forwards, power forwards, and centers tend to have higher rebound rates.

Our first step with individual-level mixture modeling for these data is to identify which compositional representation of continuous data is better suited for analyzing NBA player statistics. Next, we will present formulations of mixed membership and partial membership models and examine the data-generative capabilities of these models for both discrete and continuous data.

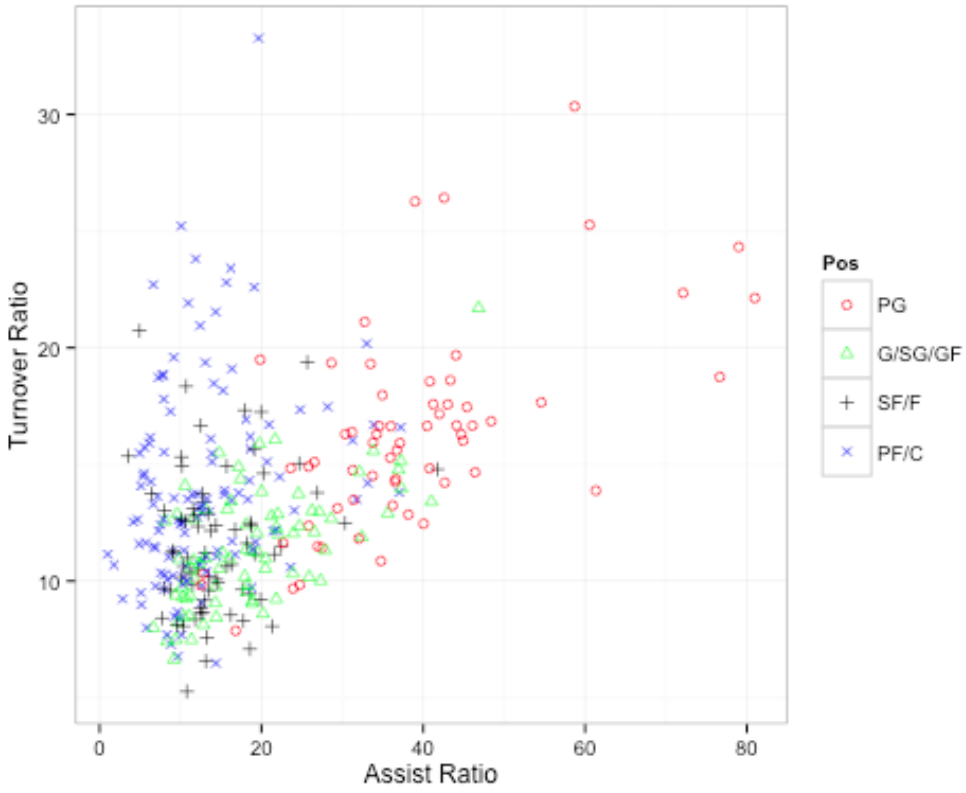


FIGURE 2.1
Bivariate scatterplot of players' assist ratio and turnover ratio. The symbols of the points represent the different positions of each player.

2.3 Two Types of Membership

In this section, we introduce the mixed membership and the partial membership models. For each of these two individual-level mixture models, we first consider a standard population-level mixture model formulation and then present each individual-level mixture model as a relaxation of the population-level mixture. Heller et al. (2008) used Bayesian methods to estimate the partial membership model. Similarly, Bayesian methods are frequently employed with mixed membership models. As we will see, the hierarchical Bayesian representations of the two models have many features in common.

2.3.1 Mixed Membership Model

Let $\mathbf{y}_i$ be a vector of p outcomes for the ith individual or observational unit. We use K to denote the number of pure types or mixture components. Let $p_k(\cdot)$ specify the density particular to pure type k, and let $\boldsymbol{\theta}_k$ represent the parameters characterizing $p_k(\cdot)$ for pure type k. The population-level mixture model with K components assumes the existence of K membership indicator variables, π_{ik}, for each individual i that designate the cluster or pure type to which the individual belongs.

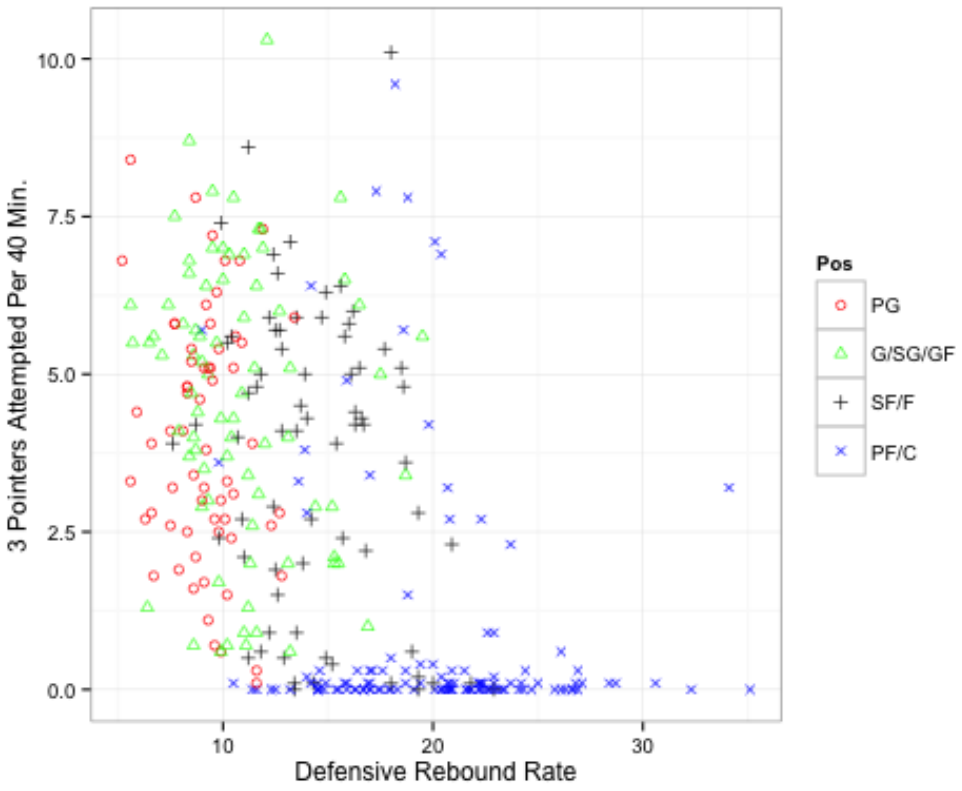


FIGURE 2.2
Bivariate scatterplot of players' defensive rebound rate and 3-point field goals attempted per 40 minutes. The symbols of the points represent the different positions of each player.

As such, $\pi_{ik} \in \{0, 1\}$ with the restriction $\sum_k \pi_{ik} = 1$. The probability density for $\mathbf{y}_i$, given a collection of parameters $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$ for all K pure types and given the latent pure type membership indicator π_{ik} for pure type k and individual i, is

$$p(\mathbf{y}_i|\boldsymbol{\Theta}, \boldsymbol{\pi}_i) = \sum_k^K \pi_{ik} p_k(\mathbf{y}_i|\boldsymbol{\theta}_k). \tag{2.1}$$

For the mixed membership model, one replaces π_{ik} with a membership score g_{ik}. Instead of being restricted to either 0 or 1, the membership score g_{ik} is allowed to range continuously between 0 and 1, subject to the constraint $\sum_k g_{ik} = 1$. The mixed membership then takes the form

$$p(\mathbf{y}_i|\boldsymbol{\Theta}, \mathbf{g}_i) = \prod_j^J \sum_k^K g_{ik} p_{jk}(y_{ij}|\theta_{jk}). \tag{2.2}$$

Here, conditional on the membership vector $\mathbf{g}_i = (g_{i1}, \ldots, g_{iK})$, the observations $\mathbf{y}_i$ are assumed to be independent.

In the Bayesian representation of the model introduced, $\mathbf{g}_i \sim D_g(\alpha, \boldsymbol{\rho})$, where D_g is a prior suitable for compostional parameters and $\alpha, \boldsymbol{\rho}$ are hyperparameters. As $\mathbf{g}_i$ lies in the $K - 1$ probability

simplex, the most common choices for D_g are the Dirichlet (Blei et al., 2003) and logistic normal (Blei and Lafferty, 2007) distributions. For the class-specific and outcome-specific parameters, θ_{jk}, a conjugate prior, is typically assumed:

$$\theta_{jk} \sim \text{Conj}(\boldsymbol{\lambda}, \boldsymbol{\nu}), \tag{2.3}$$

where $\boldsymbol{\lambda}, \boldsymbol{\nu}$ are hyperparameters. The mixed membership model has a latent class representation that suggests a data augmentation approach for estimation (Erosheva, 2003). This approach adds an additional level of hierarchy to the model by including latent classification variables.

2.3.2 Partial Membership Model

An alternative means of specifying Equation (2.1) is through the product of the densities:

$$p(\mathbf{y}_i|\boldsymbol{\Theta}, \boldsymbol{\pi}) = \prod_k^K p_k(\mathbf{y}_i|\boldsymbol{\theta}_k)^{\pi_{ik}}. \tag{2.4}$$

We specify the partial membership model by relaxing Equation (2.4) so that

$$p(\mathbf{y}_i|\boldsymbol{\Theta}, \mathbf{g}) = \frac{1}{c}\prod_k^K p_k(\mathbf{y}_i|\boldsymbol{\theta}_k)^{g_{ik}}, \tag{2.5}$$

where $g_{ik} \in [0, 1]$ and c is a normalizing constant. Heller et al. (2008) further highlights the case where p_k is an exponential family density (denoted $\text{Exp}(\cdot)$):

$$p_k(\mathbf{y}_i|\boldsymbol{\psi}_k) = \text{Exp}(\boldsymbol{\psi}_k). \tag{2.6}$$

Here, $\boldsymbol{\psi}_k$ denotes the natural parameters for pure type k. Let $\boldsymbol{\Psi}$ denote the collection of the natural parameters for all pure types.

Substituting exponential family densities for p_k in Equation (2.5), we obtain

$$p(\mathbf{y}_i|\boldsymbol{\Psi}, \mathbf{g}) = \text{Exp}\left(\sum_k g_k \boldsymbol{\psi}_k\right). \tag{2.7}$$

In addition, let the natural parameters for each pure type follow a conjugate prior distribution

$$\boldsymbol{\psi}_k \sim \text{Conj}(\boldsymbol{\lambda}, \boldsymbol{\nu}), \tag{2.8}$$

where $\boldsymbol{\lambda}, \boldsymbol{\nu}$ are hyperparameters. As with the mixed membership model, we assume $\mathbf{g}_i \sim D_g(\alpha, \boldsymbol{\rho})$ where D_g is a prior suitable for compostional parameters and $\alpha, \boldsymbol{\rho}$ are hyperparameters.

Conditional on the membership scores, $\mathbf{y}_i$ is distributed according to the same exponential family distribution as the pure types but with natural parameters that are a convex combination of the natural parameters of the pure type distributions. The use of the exponential family distributions allows one to model a variety of outcome types. Going forward, we focus on this particular case of the Bayesian partial membership model.

2.4 Comparison of Partial and Mixed Membership

In this section, we compare and contrast the partial membership model with the mixed membership model using simulated data. Figure 2.3 provides a graphical comparison of the models' data

generative processes. Although the generative structures are otherwise very similar, we see that whereas the mixed membership model assumes local independence (i.e., the outcomes are conditionally independent given the pure type memberships), the partial membership model makes no such assumption. What we can not see in Figure 2.3 is how the pure type parameters and membership scores are combined together mathematically to define the individual-level distributions of the outcomes. To explore this, we examine scatterplots for data generated by the two types of models when the data are continuous and probabilities of success generated by the respective models when the data are binary. Understanding the differences in the continuous data case will help us select an appropriate model for the NBA player data introduced in Section 2.2.

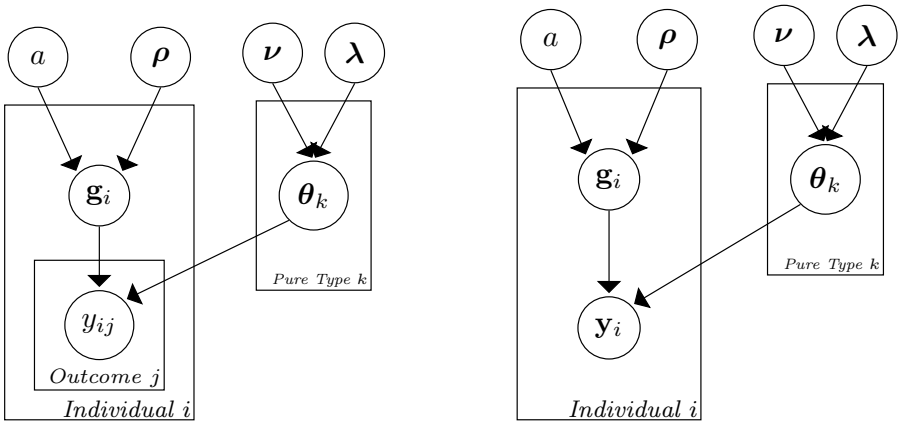


FIGURE 2.3
Graphical representations of the mixed membership (left) and partial membership models (right).

2.4.1 Continuous Data

For both the partial and mixed membership models, we begin by assuming that the pure type densities are normal. We allow the means of the normal distributions to vary by pure type. Similar to model-based clustering with the mixtures of normals (Fraley et al., 2012), different specifications are possible for the variances. For this work, we will focus on two cases. In the first case, we consider variance specifications to be the same across the pure types. In the second case, we allow the variances to differ across pure types. While the mixed membership model uses a local independence assumption, the partial membership model does not. Hence, for the partial membership model, we additionally consider two cases of variance specification: the case where the outcomes are correlated, conditional on the membership scores, and the case where the outcomes are uncorrelated, conditional on the membership scores. Next, we specify mixed membership and partial membership models for continuous data with normally distributed pure types before examining scatterplots of simulated data under the different scenarios of variance specification.

Mixed Membership

Under the mixed membership model and a local independence assumption, each outcome y_{ij}, conditional on the pure type memberships for individual i, is distributed

$$y_{ij}|\mathbf{g}_i, \boldsymbol{\Theta} \sim \sum_k g_{ik} \mathrm{N}\left(\mu_{jk}, \sigma^2_{jk}\right). \tag{2.9}$$

If we restrict $\sigma_{jk}^2 = \sigma_j^2$ so that the variances do not differ by pure type, then the model formulation remains the same. As we will see, in the case of the partial membership model, the model formulation can be simplified under the same restriction.

Partial Membership

In the case of the partial membership model, we may assume multivariate normal densities as we are not restricted to the conditional independence assumption. The observed data for individual i, $\mathbf{y}_i$ will also be multivariate normally distributed conditional on the pure type membership for individual i and the pure type parameters (recall Equation 2.7). The natural parameters of a multivariate normal distribution are $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $-\frac{1}{2}\boldsymbol{\Sigma}^{-1}$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix of a multivariate normal distribution. Let $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, k = 1, \ldots, K\}$ denote the collection of pure type means and covariance matrices. As a result, the natural parameters of $p(\mathbf{y}_i|\mathbf{g}_i, \boldsymbol{\Theta})$ are $\sum_k g_{ik}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k$ and $-\frac{1}{2}\sum_k g_{ik}\boldsymbol{\Sigma}_k^{-1}$.

Using the standard parameterization for the multivariate normal distribution, the vector of observed data, $\mathbf{y}_i$, is conditionally distributed

$$\mathbf{y}_i|\mathbf{g}_i, \boldsymbol{\Theta} \sim \mathrm{N}\left(\left(\sum_k g_{ik}\boldsymbol{\Sigma}_k^{-1}\right)^{-1}\left(\sum_k g_{ik}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k\right), \left(\sum_k g_{ik}\boldsymbol{\Sigma}_k^{-1}\right)^{-1}\right). \tag{2.10}$$

If we restrict $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_K = \boldsymbol{\Sigma}$, then

$$\mathbf{y}_i|\mathbf{g}_i, \boldsymbol{\Theta} \sim \mathrm{N}\left(\sum_k g_{ik}\boldsymbol{\mu}_k, \boldsymbol{\Sigma}\right). \tag{2.11}$$

Finally, if we assume the outcomes $\mathbf{y}_i$ are conditionally independent given the pure type memberships (local independence), each outcome y_{ij} conditional on the pure type memberships for individual i is distributed

$$y_{ij}|\mathbf{g}_i, \boldsymbol{\Theta} \sim \mathrm{N}\left(\left(\sum_k g_{ik}\sigma_{jk}^{-2}\right)^{-1}\left(\sum_k g_{ik}\sigma_{jk}^{-2}\mu_{jk}\right), \left(\sum_k g_{ik}\sigma_{jk}^{-2}\right)^{-1}\right), \tag{2.12}$$

where σ_{jk}^2 is the j-th diagonal element of $\boldsymbol{\Sigma}_k$, now a diagonal matrix.

Simulated Data Scenarios

We now compare data generated by each of the two models. Consider three pure types with two normally distributed outcomes. We present the means for each pure type in Table 2.2.

For the variance specifications, we explore two scenarios, one where the variances for each outcome are the same across pure types and a second where the variances differ across pure types. For each scenario, we consider three models: a mixed membership with a local independence assumption, a partial membership with a local independence assumption, and a partial membership model with no restrictions on dependence.

Table 2.3 summarizes Scenario 1 for which we assume the variance for the first outcome is 4 for all pure types and 9 for the second outcome for all pure types. Because of the local independence assumption used in the mixed membership model, there is no correlation between the two outcomes. As a means of comparison, we consider a corresponding partial membership model that employs the local independence assumption and hence also has the correlation between the two outcomes restricted to 0. Finally, the partial membership model without a local independence assumption assumes a correlation of 0.4.

Table 2.4 presents the corresponding information for Scenario 2 where the covariances may vary by pure type. For the partial membership model with full covariance matrix, the correlations by pure type were set to 0.4, -0.4, and 0.7.

TABLE 2.2
Pure Type Means.

	Pure Type		
Outcome	**1**	**2**	**3**
1	10	25	40
2	25	40	10

TABLE 2.3
Covariance matrices under Scenario 1.

Model	**Pure Types 1–3**
Mixed Membership	$\begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}$
Partial Membership (Uncorrelated)	$\begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}$
Partial Membership (Correlated)	$\begin{pmatrix} 4 & 2.4 \\ 2.4 & 9 \end{pmatrix}$

TABLE 2.4
Covariance matrices under Scenario 2.

	Pure Type		
Model	**1**	**2**	**3**
Mixed Membership	$\begin{pmatrix} 4 & 0 \\ 0 & 16 \end{pmatrix}$	$\begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}$
Partial Membership (Uncorrelated)	$\begin{pmatrix} 4 & 0 \\ 0 & 16 \end{pmatrix}$	$\begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}$
Partial Membership (Correlated)	$\begin{pmatrix} 4 & 3.2 \\ 3.2 & 16 \end{pmatrix}$	$\begin{pmatrix} 9 & -1.2 \\ -1.2 & 1 \end{pmatrix}$	$\begin{pmatrix} 9 & 4.2 \\ 4.2 & 4 \end{pmatrix}$

Scenario 1: Same Variances Across Pure Types

Under Scenario 1, we assume the pure type covariances are common across all three pure types. We keep the population parameters constant and vary the distribution of membership scores to produce scatterplots of observed data.

We generated 1000 random membership vectors from a Dirichlet ($a\boldsymbol{\rho}$) distribution with $a = 1$ and $\boldsymbol{\rho} = (1/3, 1/3, 1/3)$. Using these membership scores, we simulated 1000 bivariate outcomes. The results are depicted in Figure 2.4(b). The left plot shows the mixed membership model and the center plot displays the corresponding partial membership model with a diagonal covariance matrix (i.e., local independence was assumed as in the case of the mixed membership model). The right plot shows partial membership model results with a full covariance matrix where the variances of the outcomes are the same as the previous two cases but the correlation between the outcomes is set to 0.4.

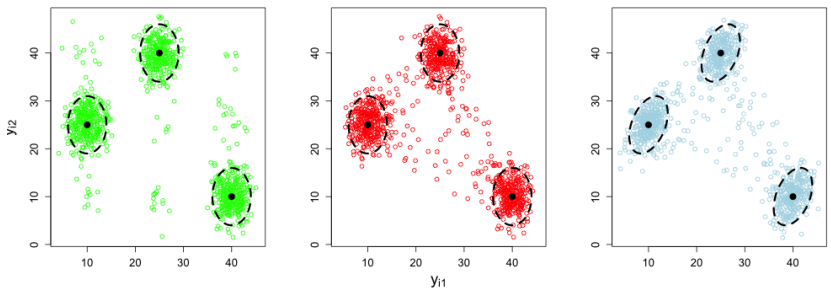
In Figure 2.4(b), the mixed membership model generates points in three columns. Looking more closely, each column can be divided horizontally into three parts corresponding to the means for each pure type for y_{i2}. Dividing the columns in this manner produces $K^2 = 9$ clusters of points, consistent with the latent class representation described by Erosheva (2006) and the more extreme depiction presented in Figure 4 in Heller et al. (2008). The partial membership model, in both the diagonal and full covariance matrix cases, generates points in a more cloud-like structure. One can see that the partial membership model with the full covariance matrix generates a set of points that is "rotated," albeit slightly, as compared to the set generated by the partial membership model with a diagonal covariance matrix.

By varying the values of a, we can further compare the models. If we set $a = 10$, the membership scores will fluctuate more closely around 1/3 than $a = 1$. Figure 2.4(c) presents 1000 generated data points with membership scores generated from a Dirichlet ($a\boldsymbol{\rho}$) distribution with $a = 10$ and $\boldsymbol{\rho} = (1/3, 1/3, 1/3)$. In the case of the mixed membership model, the K^2 clusters become slightly more apparent while the data generated by the partial membership models reduce to single clusters with less variation. If we set $a = 1/10$, the membership scores tend to be closer to the extremes 0 or 1. Figure 2.4(a) presents the simulated data from each model with this set of membership scores. The three plots now appear largely similar. The primary differences are that the set of points generated by the partial membership model with full covariance matrix is "rotated" as compared to the other two and that the mixed membership model appears to show greater variation in points on the periphery.

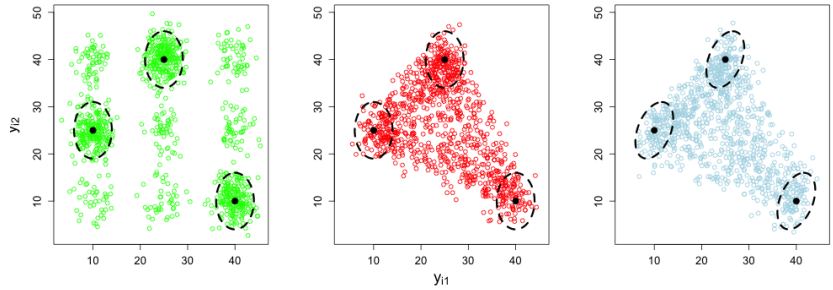
Scenario 2: Different Variances Across Pure Types

We subsequently generate data points from each individual-level mixture model according to Scenario 2. Again, the pure type covariances for Scenario 2 are listed in Table 2.4. Figure 2.5(b) presents the data generated by the mixed membership model, the partial membership model with diagonal covariance, and the partial membership model with unrestricted covariance for $a = 1$. The sets of points generated by the mixed membership and partial membership model with diagonal covariance appear rectangular in shape. The set of points from the partial membership model with diagonal covariance is more densely populated in the center while one can faintly make the clusters in the set of points generated by the mixed membership model. The partial membership model with full covariance matrices on the other hand is more triangular in structure.

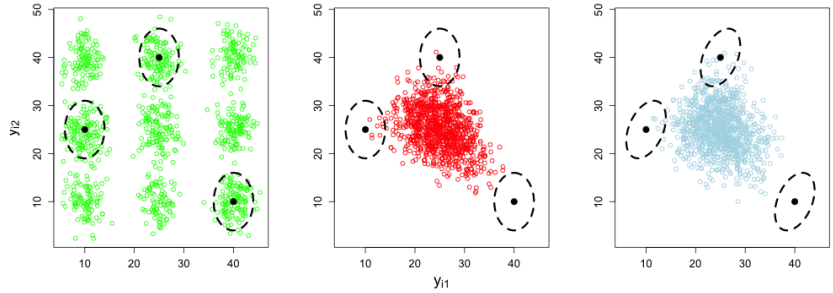
Figures 2.5(c) and 2.5(a) provide the corresponding plots for membership vectors generated by $a = 10$ and $a = 1/10$, respectively. With $a = 10$, we again see the greater concentration of points into a single cluster for the partial membership models while the different clusters become a little more apparent for the mixed membership model. In the case of $a = 1/10$, the mixed membership and partial membership with diagonal covariance models again appear very similar. The full covariance partial membership model, however, displays a triangular boundary with an empty center.



(a) $a = 0.1$

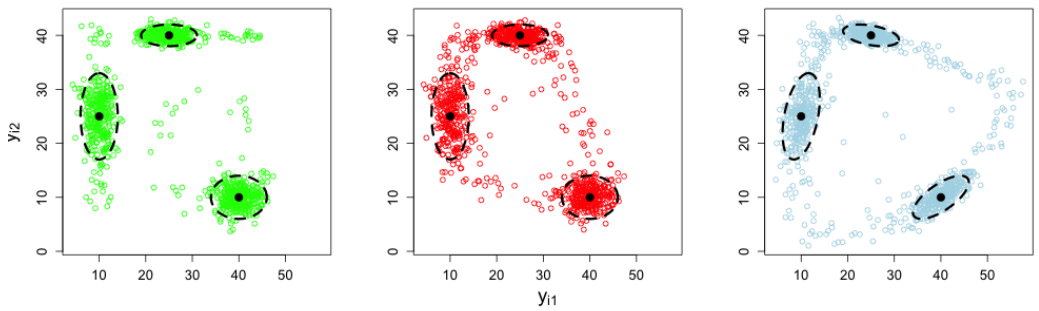


(b) $a = 1$

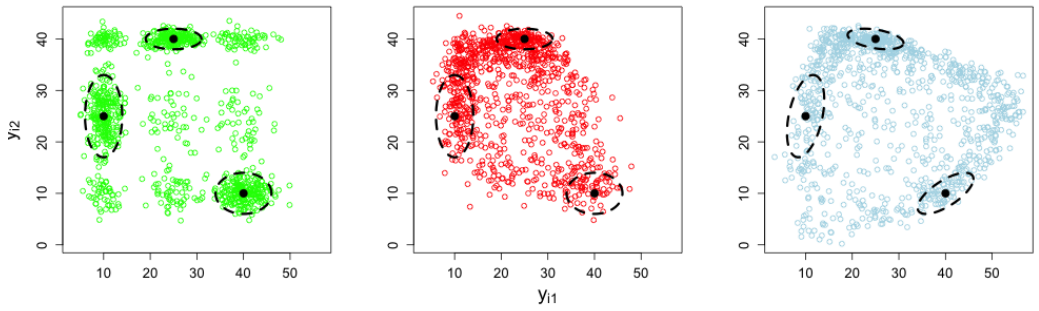


(c) $a = 10$

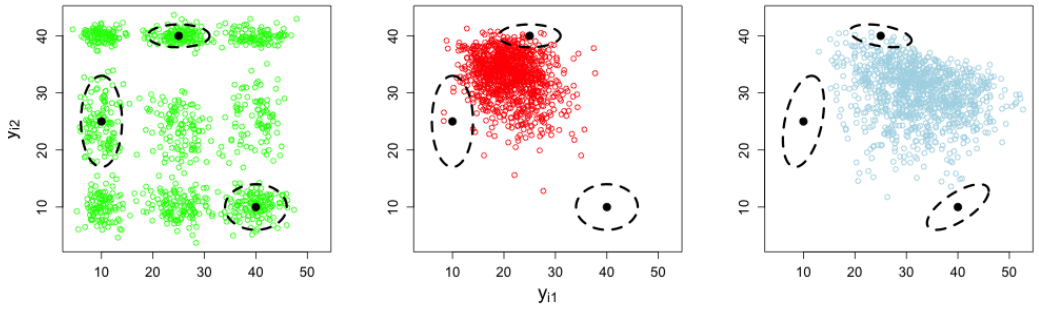
FIGURE 2.4
Simulated data according to different individual-level mixture models assuming variances are the same across pure types. Each panel contains: mixed membership (left), partial membership with local independence assumption (center), partial membership with full covariance matrix (right). The solid points represent the pure type centers and the dashed ellipses represent 2SD contours.



(a) $a - 0.1$



(b) $a = 1$



(c) $a = 10$

FIGURE 2.5
Simulated data according to different individual-level mixture models for the case where the variances are different across pure types. Each panel contains: mixed membership (left), partial membership with local independence assumption (center), partial membership with full covariance matrix (right). The solid points represent the pure type centers and the dashed ellipses represent 2SD contours.

Overall, while the mixed and partial membership models can produce scatterplots that look very similar for some special cases of the distribution of the membership scores, we observe that the partial membership models generate scatterplots that are more contiguous. At the same time, we emphasize that placements of pure type means and variances, as well as the selection of the distribution of the membership scores, can create different patterns that would not be as easy to recognize as either mixed or partial membership. To investigate this further, one could consider a template for variance specification as provided by model-based clustering with Gaussian clusters (Fraley et al., 2012).

2.4.2 Binary Data

We now examine the mixed and partial membership models for binary data. We follow a geometric approach (Erosheva, 2005) where we keep the population parameters constant and examine population heterogeneity manifolds obtained by letting subject-level parameters vary over their natural range.

In the case of binary data, we compare the models by examining the probability of a positive response, $p(y_{ij} = 1|\mathbf{g}_i, \mathbf{\Theta})$, for outcome j and individual i, conditional on the pure type membership of individual i. Let θ_{jk} denote the probability of a positive response for pure type k and outcome j. Then,

$$\theta_{ij} = p(y_{ij} = 1|\mathbf{g}, \mathbf{\Theta}) = \sum_k g_{ik}\theta_{jk}, \tag{2.13}$$

so that $y_{ij}|\mathbf{g}, \mathbf{\Theta}$ has a Bernoulli distribution where the probability of a positive response is a weighted arithmetic mean of the pure type response probabilities.

In the case of the partial membership model, $y_{ij}|\mathbf{g}, \mathbf{\Theta}$ also has a Bernoulli distribution but where the natural parameter is a convex combination of the pure type natural parameters, $\sum_k g_{ik}\ln[\theta_{jk}/(1-\theta_{jk})]$. As a result,

$$\theta_{ij} = p(y_{ij} = 1|\mathbf{g}, \mathbf{\Theta}) = \frac{\prod_k \theta_{jk}^{g_{ik}}}{\prod_k \theta_{jk}^{g_{ik}} + \prod_k (1-\theta_{jk})^{g_{ik}}}. \tag{2.14}$$

In the case of the partial membership model, the probability of a positive response (Equation 2.14) is a normalized weighted geometric mean of the pure type response probabilities.

We now examine how these differences in the mixed membership and partial membership models for binary data manifest themselves for different pure type membership and parameter values. We consider $K = 2$ pure types and $p = 2$ outcomes. Let g_i denote the degree of membership for an arbitrary individual in the first pure type; the degree of membership in the second pure type is then $1 - g_i$. We examine θ_{ij}, the marginal probability of a positive response for outcome j, and individual i given by Equations (2.13) and (2.14) for the two types of models, respectively.

Table 2.5 presents five sets of the pure type response probabilities, θ_{jk}; the corresponding marginal probability plots appear in Figure 2.6. Treating the pure type response probabilities as constant, we examine population heterogeneity manifolds obtained by letting membership scores g_i vary over their natural range from 0 to 1. The darker points indicate the population heterogeneity manifolds obtained with the partial membership model for given θ_{i1} and θ_{i2}, whereas the lighter points indicate the corresponding manifolds for the mixed membership model.

For Scenario 1, we see that the heterogeneity manifold for the partial membership model is a nonlinear path that closely resembles the heterogeneity manifold for the mixed membership model. As the pure type response probability θ_{11} decreases and θ_{22} increases over the five scenarios, the paths of points increasingly diverge. Finally, for Scenario 5, the partial membership model produces the heterogeneity manifold that takes only three pairs of values, sitting at the corners of the marginal probability space. At $g_i = 0$ and $g_i = 1$, the partial membership model produces θ_{ij} values

TABLE 2.5
Pure type response probabilities.

Scenario	θ_{11}	θ_{12}	θ_{21}	θ_{22}
1	0.1	0.8	0.3	0.6
2	0.05	0.8	0.3	0.95
3	0.01	0.8	0.3	0.99
4	0.001	0.8	0.3	0.999
5	0	0.8	0.3	1

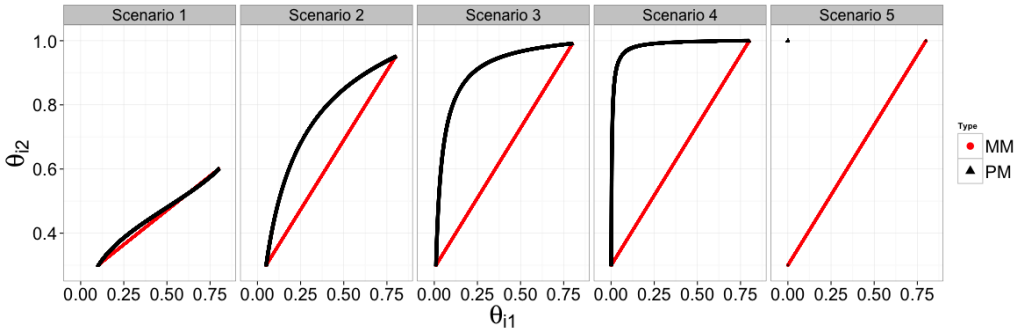


FIGURE 2.6
Marginal probability plots for Scenarios 1–5 in Table 2.5 obtained with the partial membership model (darker) and the mixed membership model (lighter).

equivalent to the mixed membership model. For values of g_i, $0 < g_i < 1$, in Scenario 5, $\theta_{i1} = 0$, and $\theta_{i2} = 1$ under the partial membership model. Consistent with the geometric mean representation in Equation (2.14), in scenarios where one of the pure type conditional response probabilities equals 1, any partial membership in the pure type implies that that individual's probability for that outcome must be 1. Similarly, when one of the pure type conditional response probabilities equals 0, any partial membership in that pure type implies that the probability for that outcome must be 0. We do not observe this property in the mixed membership model that employs the arithmetic mean to derive individual-specific marginal probabilities (Equation 2.13). Moreover, as one of the pure type probabilities decreases to 0 or increases to 1, the population heterogeneity manifolds obtained under the partial membership and mixed membership models increasingly diverge, as shown in Figure 2.6.

Overall, we have demonstrated that the partial and mixed membership models exhibit different data-generating behavior. In the case of continuous data, the partial membership model generates data in more contiguous patterns that may be more natural for some applications. However, except for some special cases, it may not be possible to tell the nature of individual-level mixing from scatterplots. Hence, data mechanisms need to be considered.

Our decision for the analysis of the NBA data is to use a partial membership model. We believe that a partial membership model could better describe the types of data patterns displayed in Figures 2.1 and 2.2 than a mixed membership model. However, an equally important factor in our decision is the nature of individual-level mixing in the data. The NBA player data contain variables that themselves are summary statistics as opposed to individual player's actions. While mixed membership modeling should be more appropriate for the latter type of data that could exhibit changes in (latent) pure type assignments for each variable, we find the partial membership representation to

be more consistent with the averages reported over an NBA season. These considerations are akin to the switching and blending interpretations discussed in Galyardt (2014).

2.5 A Correlated Partial Membership Model for Continuous Data

Before analyzing the NBA player style data with a partial membership model for continuous data, we develop an extension of the partial membership model that allows for correlated membership scores. We subsequently discuss estimation of the correlated partial membership model.

2.5.1 Correlated Memberships

One limitation of the partial membership model as originally formulated is its inability to flexibly accommodate correlations among an individual's membership in the pure types. The Dirichlet prior induces a small negative correlation among the pure type memberships in individuals. Blei and Lafferty (2007) addressed this shortcoming in mixed membership topic models by replacing the Dirichlet prior for individual membership scores with a logistic normal prior. Under this model, draws from the multivariate normal are transformed to map the probability simplex so that the values are positive and constrained to add to 1,

$$\boldsymbol{\eta}_{\mathbf{g}_i} \sim \mathrm{N}\left(\boldsymbol{\rho}, \boldsymbol{\Sigma}\right), \tag{2.15}$$

$$g_{ik} = \frac{\exp(\eta_{g_{ik}})}{\sum_l \exp(\eta_{g_{il}})}. \tag{2.16}$$

Because of the constraints that $\sum_k g_{ik} = 1$, we fix the Kth element of $\boldsymbol{\eta}_{\mathbf{g}_i}$ to 0 so that the vector contains only $K-1$ free elements and $\boldsymbol{\rho}$ and $\boldsymbol{\Sigma}$ have dimensions $K-1$ and $(K-1)\times(K-1)$, respectively. Atchison and Shen (1980) discuss properties and uses of the logistic normal, including a comparison with the Dirichlet distribution. They suggest that the logistic normal can suitably approximate the Dirichlet distribution so that little, if anything, would be lost if we applied the logistic normal in cases where a Dirichlet prior would be appropriate.

2.5.2 A Correlated Partial Membership Model

To model the continuous data in the NBA example, we assume the observed data points for individual i, $\mathbf{y}_i$ are conditionally independent given the pure type memberships for the individual, $\mathbf{g}_i$. Equation (2.9) gives the distribution of y_{ij} under this assumptions. Now let $\tau_{jk} = \sigma_{jk}^{-2}$ and let $\alpha_{jk} = \sigma_{jk}^{-2}\mu_{jk}$ in Equation (2.9) so that τ_{jk} and ϕ_{jk} correspond closely to the natural parameters of a normal distribution. Moreover, let $\Theta = \{g_{ik}, \phi_{jk}, \tau_{jk}, \rho_k, j = 1, \ldots, J, k = 1, \ldots, K\}$.

For α_{jk} and τ_{jk}, we specify normal and gamma prior distributions, respectively. The elements of the mean vector of the untransformed pure type memberships, ρ_k, are also specified to have normal prior distributions. For the covariance matrix for the untransformed pure type memberships, $\boldsymbol{\Sigma}$, we use an inverse Wishart prior distribution. Fully stated, the correlated partial membership model for continuous data is

$$y_{ij}|\mathbf{g}_i, \boldsymbol{\Theta} \sim \mathrm{N}\left(\left(\sum_k g_{ik}\tau_{jk}\right)^{-1}\left(\sum_k g_{ik}\alpha_{jk}\right), \left(\sum_k g_{ik}\tau_{jk}\right)^{-1}\right), \tag{2.17}$$

$$\alpha_{jk} \sim \mathrm{N}\left(m_{\alpha_{jk}}, s^2_{\alpha_{jk}}\right), \tag{2.18}$$

$$\tau_{jk} \sim \text{Gamma}\left(\nu_{\tau_{jk}}, \phi_{\tau_{jk}}\right), \tag{2.19}$$

$$g_{ik} = \frac{\exp(\eta_{g_{ik}})}{\sum_l \exp(\eta_{g_{il}})}, \tag{2.20}$$

$$\boldsymbol{\eta}_{\mathbf{g}_i} \sim \text{N}\left(\boldsymbol{\rho}, \boldsymbol{\Sigma}\right), \tag{2.21}$$

$$\rho_k \sim \text{N}\left(m_{\rho_k}, s^2_{\rho_k}\right), \tag{2.22}$$

$$\boldsymbol{\Sigma} \sim \text{Inv. Wishart}\left(\nu_\Sigma, S_\Sigma\right). \tag{2.23}$$

In order to obtain posterior samples of μ_{jk} and σ^2_{jk} rather than α_{jk} and τ_{jk}, we may transform the posterior samples of α_{jk} and τ_{jk} to μ_{jk} and σ^2_{jk}.

2.5.3 Estimation

Let $\boldsymbol{\Omega}$ denote the set of hyperparameters for the prior distributions of the parameters in $\boldsymbol{\Theta}$ as specified in Equation (2.17). The joint probability of $\mathbf{Y}$ and $\boldsymbol{\Theta}$ conditional upon $\boldsymbol{\Omega}, \boldsymbol{\Sigma}$ is

$$\begin{aligned}
p\left(\mathbf{Y}, \boldsymbol{\Theta} | \boldsymbol{\Sigma}, \boldsymbol{\Omega}\right) = & \prod_i^I \prod_j^J \left(2\pi\right)^{-1/2} \left(\sum_k^K g_{ik}\tau_{jk}\right)^{1/2} \\
& \cdot \exp\left(-\frac{\sum_k^K g_{ik}\tau_{jk}}{2}\left[y_{ij} - \left(\sum_k^K g_{ik}\tau_{jk}\right)^{-1} \sum_k^K g_{ik}\alpha_{jk}\right]^2\right) \\
& \prod_j^J \prod_k^K \left(2\pi s^2_{\alpha_{jk}}\right)^{-1/2} \exp\left(-\frac{1}{2s^2_{\alpha_{jk}}}\left(\alpha_{jk} - m_{\alpha_{jk}}\right)^2\right) \\
& \prod_j^J \prod_k^K \frac{\phi_{\tau_{jk}}^{\nu_{\tau_{jk}}}}{\Gamma\left(\nu_{\tau_{jk}}\right)} \tau_{jk}^{\nu_{\tau_{jk}}-1} \exp\left(-\phi_{\tau_{jk}}\tau_{jk}\right) \\
& \prod_i^I \left(2\pi\right)^{-(K-1)/2} \left|\boldsymbol{\Sigma}\right|^{-1/2} \exp\left(-\frac{1}{2}\left(\boldsymbol{\eta}_i - \rho\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\eta}_i - \rho\right)\right) \\
& \prod_k^{K-1} \left(2\pi s^2_{\rho_k}\right)^{-1/2} \exp\left(-\frac{1}{2s^2_{\rho_k}}\left(\rho_k - m_{\rho_k}\right)^2\right).
\end{aligned} \tag{2.24}$$

As Heller et al. (2008) noted, all of the parameters in $\boldsymbol{\Theta}$ are continuous and, moreover, we may take the derivatives of the log of the above probability expression. As a result, the problem of Bayesian estimation for this model lends itself to Hybrid (Hamiltonian) Monte Carlo. Hybrid Monte Carlo uses the derivative of the log joint probability to inform its proposals. As a result, in high dimensions, this algorithm may outperform more traditional algorithms such as Metropolis-Hastings or Gibbs sampling. For a thorough introduction to Hybrid Monte Carlo, see Neal (2010). In order to avoid the imposition of non-negativity restrictions on τ_{jk} in the Hybrid Monte Carlo algorithm, we employ the transformation $\eta_{\tau_{jk}} = \log(\tau_{jk})$ so that the parameter may take values unrestricted over the real line.

We do not rely on Hybrid Monte Carlo to draw $\boldsymbol{\Sigma}$ but rather draw $\boldsymbol{\Sigma}$ in a separate Gibbs step for the correlated partial membership model. Thus, to sample $(\boldsymbol{\Theta}, \boldsymbol{\Sigma})$, we apply a Gibbs sampling algorithm where the first step involves sampling $\boldsymbol{\Theta}$ via Hybrid Monte Carlo and then $\boldsymbol{\Sigma}$ from its full conditional distribution,

$$\boldsymbol{\Sigma} \sim \text{Inv. Wishart}\left(\nu_{\boldsymbol{\Sigma}} + n, S_{\boldsymbol{\Sigma}} + \left(\mathbf{H}_G - \mathbf{1}_n\boldsymbol{\rho}^T\right)^T\left(\mathbf{H}_G - \mathbf{1}_n\boldsymbol{\rho}^T\right)\right), \tag{2.25}$$

where $\mathbf{H}_G$ is a $n \times K-1$ matrix of the untransformed membership scores.

2.6 Application to the NBA Player Data

We now apply the correlated membership model to NBA player data from the 2010–11 season. We considered models with 4, 5, and 6 pure types. We employed posterior predictive model checks to examine the fit of the model-based marginal distributions and rank correlations to the observed data. We ultimately settled on a model with 5 pure types as this model had the smallest number of classes that still provided sufficient fit to the data. The 5 pure type correlated partial membership model resulted in easily interpretable classes from a substantive viewpoint. Also, each pure type had at least one membership score above 0.20, meaning that at least one player had 1/5 or more of their membership in that type.

We ran the Gibbs sampling algorithm with a Hybrid Monte Carlo step for 80,000 iterations, keeping every 20th draw. We discarded the first 1000 of the retained draws as burn-in, leaving us with 3000 samples from the posterior distribution. To asses convergence, we examined trace plots and used the Geweke (Geweke, 1992) and Raftery-Lewis (Raftery and Lewis, 1995) diagnostic tests.

In examining the posterior estimates for the pure type specific means, μ_{jk}, presented in Table 2.6, we notice that some of the posterior means take negative values when all of the statistics recorded are strictly positive. For example, in the case of the % Ast statistic (the percentage of made field goals that are assisted), the range of the data is [0, 100], yet only one of the estimated pure type means lies inside this range. This observation is not worrisome by itself as it could be that no individual has high membership values in the pure types with negative means. We are more concerned with the associated predictive distributions for the observed data that are directly related to model fit. Nonetheless, when a pure type is characterized by values outside the range of observed data, the interpretation of this pure type is more complicated than of those pure types that can in principle be achievable in the population.

Figure 2.7 presents a posterior predictive model check that compares the marginal distribution of the percent of made fields goals assisted (% Ast) statistic against the replicated values for the statistic. The histogram depicts the observed data while the black points represent the posterior predictive mean count of replicated values falling in the corresponding bin. The black segment represents the 95% credible interval. We observe in Figure 2.7 that the model fits the marginal distribution of the data well; we obtained similar findings for other variables (not shown).

Although the model provides a good fit to the observed data, the shortcoming of this model is that it still places (small) non-zero predictive density in the improbable region of the data. This shortcoming will naturally arise when we use a normal distribution to model range-restriced data.

Examining the ordering of the posterior means can provide us with a way to characterize the pure types in relation to one another. Table 2.6 illustrates that pure type 1 comprises players who play a high number of minutes (Min), have a high percentage of their shots assisted (% Ast), shoot mid- and long-range jumpers (Medium, Long, 3s), and have a low steals rate (Stls). We refer to this pure type as the "high minute shooters." A high percentage of shots assisted (% Ast) and high volume of 3-point shots (3s) also describes pure type 2, but members of this pure type have fewer shots at all other distances (Rim, Close, Medium, Long) and a lower number of minutes played. We refer to this pure type as the "3-point specialists." The posterior means for the 3rd pure type are high relative to those for the other pure types across almost all variables except for the the mid- to long-range jumpers (Medium, Long, 3s). We use the term "active player" for this pure type. Low minutes played (Min) and high offensive rebound rates (ORR) are the most distinguishing features of pure type 4 which we refer to as the "limited big men" pure type. High assist (AR) and turnover (TOR) ratios, high steals per 40 minutes, a low percentage of shots assisted and low blocked shots per 40 minutes mark the final pure type. We refer to this pure type as the "ball handlers" pure type.

Figure 2.8 presents the mean posterior memberships of the players in these different pure types. The points' symbols denote their assigned position recorded in the original dataset. Here, we can see

TABLE 2.6
Posterior means for pure type mean parameters, μ_{jk}.

	Pure Type				
Var.	**1**	**2**	**3**	**4**	**5**
Min	42.40	13.36	416.04	18.09	29.43
% Ast	132.09	108.08	-179.33	70.81	-12.47
AR	4.36	131.68	916.56	9.20	607.91
TOR	6.55	-132.22	210.90	16.96	194.49
ORR	-29.30	-5.90	1552.00	9.34	1.70
DRR	228.25	7.51	1077.20	17.26	8.01
Rim	0.26	-0.64	116.07	3.73	6.03
Close	14.73	-0.13	238.88	1.38	2.07
Medium	340.24	0.08	10.87	1.17	3.35
Long	77.21	0.72	-33.99	3.05	3.81
3s	10.80	15.21	22.61	0.02	2.14
Stls	-3.51	0.96	54.93	0.81	1.82
Blks	12.81	0.25	61.76	1.84	0.21

that high membership in some pure types corresponds to certain positional assignments. Thus, the highest memberships in the limited big men pure type (pure type 4) are obtained by centers (C) and power forwards (PF) while the ball handlers pure type (pure type 5) is dominated by point guards (PG). Membership in pure types 1–3 does not have a close correspondence with specific assigned positions. For pure types 1–3, and to a lesser extent for pure type 5, no players come close to being fully represented by the pure type. This explains why the model performs well for predicting the marginal probability for the % Ast outcome despite having posterior means for pure types 1–3 and 5 to be out of bounds on that variable.

In contrast to the original partial membership model with Dirichlet membership scores (Heller et al., 2008), the correlated partial membership model allows for a more flexible correlation structure among components of the membership vector. Table 2.7 presents the posterior mean correlations of the pure type memberships that range from -0.664 to 0.410. The limited big man pure type (pure type 4) shows low to moderate negative correlations with all other pure types. The active player pure type, on the other hand, shows low to moderate positive correlations with the high minute shooter and 3-point specialist pure types and small negative correlations with the limited big man and ball handler pure types. We note that it is impossible to observe positive correlations under the Dirichlet type models. This suggests that our decision to allow for more flexible modeling of the pure type membership correlations was appropriate for the data.

TABLE 2.7
Posterior mean correlations of membership scores.

	1	2	3	4	5
1	1.000	0.081	0.410	-0.528	0.160
2	0.081	1.000	0.223	-0.553	-0.080
3	0.410	0.223	1.000	-0.235	-0.328
4	-0.528	-0.553	-0.235	1.000	-0.664
5	0.160	-0.080	-0.328	-0.664	1.000

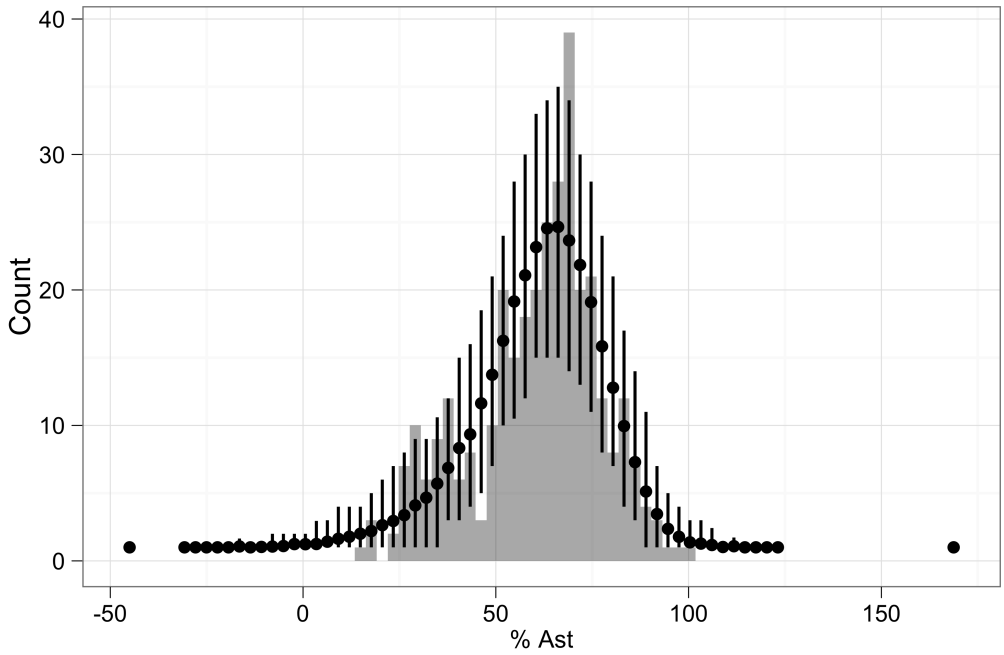


FIGURE 2.7
Histogram of the observed values for the % Ast statistic. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

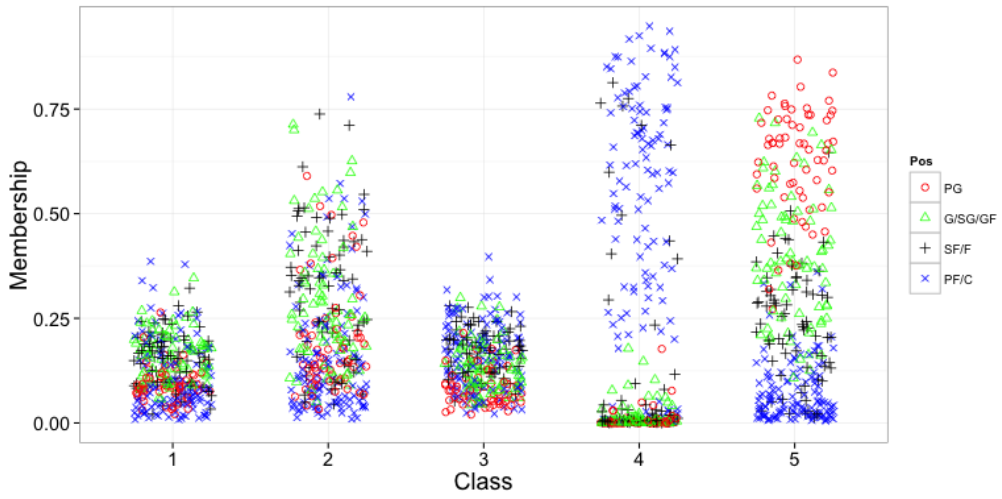


FIGURE 2.8
The mean posterior memberships of the players by pure type. The shapes of the points represent the different positions of each player.

To explore the compositional styles of NBA players further, consider the posterior mean membership scores and the corresponding credible intervals for three NBA "combo guards": Mario Chalmers, Steve Blake, and Rudy Fernandez, as identified by Lutz (2012). As a point of contrast, we examine the corresponding quantities for Chris Paul, who is generally considered to be an example of a pure point guard (Figure 2.9). We observe that 80% of Chris Paul's membership is in the ball handlers pure type. For the other three players, their membership is largely split between the ball handlers pure type and the 3-point specialists. Thus, we see that the correlated partial membership model describes the combo guard players using a mixture of pure types. This result stands in contrast to the results of the cluster analysis performed by Lutz (2012), where the combo guards comprised their own cluster, entirely separate from the other 12 clusters found in that analysis. Our correlated partial membership model uses only 5 pure types but characterizes the heterogeneity in individual playing styles as combinations of these pure types.

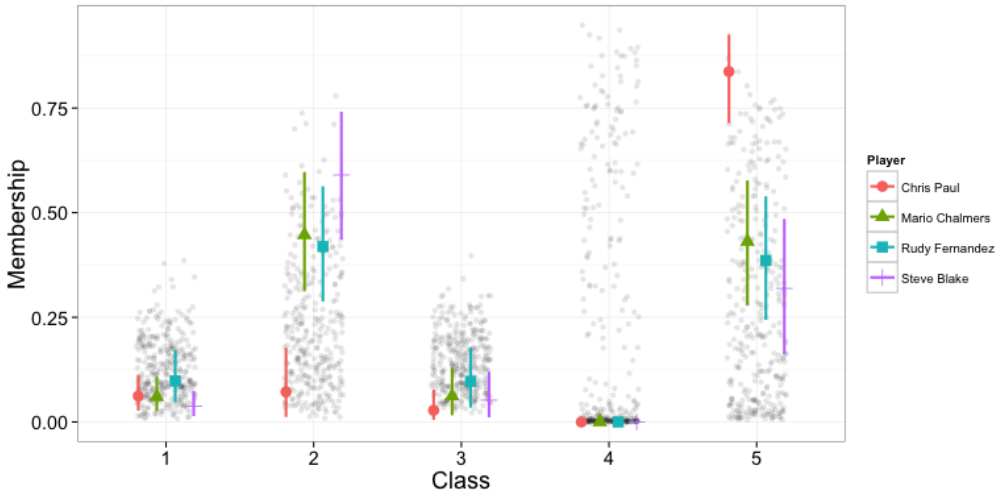


FIGURE 2.9
The mean posterior memberships and 95% posterior credible intervals of Chris Paul, Mario Chalmers, Rudy Fernandez, and Steve Blake. The grey points represent the posterior mean memberships of the other players in the data.

2.7 Summary and Discussion

In this chapter, we explored two individual-level mixture models for latent compositional data, namely, the mixed and partial memberships models. We found that the partial membership model has better potential for producing realistic representations of contiguous data patterns. However, we note that high-dimensional multivariate distributions of real data typically present even more complexity than the simulated examples considered here, which could easily mask the soft cluster-

ing nature of the underlying process. In such cases, one should consider a plausible interpretation for the latent compositional data at hand. For example, we point out that the partial membership formulation is consistent with the blending interpretation of mixed membership models as proposed by Galyardt (2014), because the NBA player dataset is primarily composed of continuous summary statistics. By contrast, in the binary data case, depending on the placement of pure type response probabilities, we observe that the partial membership model may result in a very particular behavior where fewer outcome combinations are possible compared to the Grade of Membership model. The implication of this finding for individual-level mixture models with binary data is that partial membership may not be appropriate for all binary data cases.

We modified the partial membership model to incorporate a logistic normal distribution for pure type memberships, similar to the correlated topic model extension (Blei and Lafferty, 2007) of the latent Dirichlet allocation models (Blei et al., 2003). This approach gave us more flexibility in specifying the dependence structure among the pure type memberships. We have illustrated the use of a partial membership model on continuous data using NBA player statistics. The NBA dataset provided an illustrative example where pure type membership scores exhibited both negative and positive correlations. We note that it is not possible to obtain positive correlations when one employs a Dirichlet distribution for the membership scores.

Although our partial membership analysis of the NBA player data resulted in a good fit as measured by the posterior predictive model checks, the limitation of using Gaussian pure type distributions is that the predicted values may lie outside of the allowable data intervals for variables that are constrained in their range. While it may be possible to specify other distributions for the pure types that can produce suitably constrained predicted values, a more general semiparametric approach that can accommodate not only range-restricted variables but also mixed data with both discrete and continuous outcomes could be more beneficial going forward (Gruhl et al., 2013). Examples of mixed outcome data are increasingly common in medicine and the social sciences, and the development of individual-level mixture models could be helpful for characterizing patterns in multivariate mixed outcomes.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008). Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research* 9: 1981–2014.

Atchison, J. and Shen, S. (1980). Logistic-normal distributions: Some properties and uses. *Biometrika* 67: 261–272.

Blei, D. M. and Lafferty, J. D. (2007). A correlated topic model of Science. *Annals of Applied Statistics* 1: 17–35.

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

Brustein, J. (2012). Data Crunchers Look to Quantify Chemistry in N.B.A. Off The Dribble: The New York Times N.B.A. Blog.

Erosheva, E. A. (2002). Grade of Membership and Latent Structure Models with Application to Disability Survey Data. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Erosheva, E. A. (2003). Bayesian estimation of the Grade of Membership model. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M. (eds), *Bayesian Statistics 7*. New York, NY: Oxford University Press, 501–510.

Erosheva, E. A. (2004). Partial membership models with application to disability survey data. In *Statistical Data Mining and Knowledge Discovery*. Chapman & Hall/CRC, 117–134.

Erosheva, E. A. (2005). Comparing latent structures of the Grade of Membership, Rasch, and latent class models. *Psychometrika* 70: 619–628.

Erosheva, E. A. (2006). Latent Class Representation of the Grade of Membership Model. Tech. report 492, University of Washington.

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

Fraley, C., Raftery, A. E., Murphy, T. B., and Scrucca, L. (2012). MCLUST Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Tech. report 597, University of Washington.

Galyardt, A. (2014). Interpreting mixed membership: Implications of Erosheva's representation theorem. In Airoldi, E. M., Blei, D. M., Erosheva, E. A., and Fienberg, S. E. (eds), *Handbook of Mixed Membership Models and Its Applications*. Chapman & Hall/CRC.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bernardo, J., Berger, J., Dawid, A., and Smith, J. (eds), *Bayesian Statistics 4*. Oxford, UK: Oxford University Press, 169–193.

Gormley, I. C. and Murphy, T. B. (2009). A Grade of Membership model for rank data. *Bayesian Analysis* 4: 265–296.

Gruhl, J., Erosheva, E. A., and Crane, P. (2013). A semiparametric approach to mixed outcome latent variable models: Estimating the association between cognition and regional brain volumes. *Annals of Applied Statistics* To appear.

Heller, K. A., Williamson, S., and Ghahramani, Z. (2008). Statistical models for partial membership. In *Proceedings of the 25th International Conference on Machine Learning (ICML '08)*. New York, NY, USA: ACM, 392–399.

Hoopdata (2012). NBA Player Statistics. www.hoopdata.com.

Lazarsfeld, P. F. and Neil, H. W. (1968). *Latent Structure Analysis*. Boston, MA: Houghton Mifflin.

Lutz, D. (2012). A cluster analysis of NBA players. In *Proceedings of the MIT Sloan Sports Analytics Conference*. Boston, MA, USA.

Neal, R. M. (2010). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G., and Meng, X. -L. (eds), *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC, 113–162.

Pritchard, J., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Raftery, A. E. and Lewis, S. (1995). The number of iterations, convergence diagnostics and generic metropolis algorithms. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. (eds), *Markov Chain Monte Carlo in Practice*. London, U.K.: Chapman and Hall.

Rogers, S., Girolami, M., Campbell, C., and Breitling, R. (2005). The latent process decomposition of cDNA microarray data sets. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2: 143–156.

Wang, C., Blei, D. M., and Fei-Fei, L. (2009). Simultaneous image classification and annotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*. Los Alamitos, CA, USA: IEEE Computer Society, 1903–1910.

Woodbury, M. A., Clive, J., and Garson, A., Jr. (1978). Mathematical typology: A Grade of Membership technique for obtaining disease definition. *Computers and Biomedical Research* 11: 277–298.