Published in final edited form as:

Biometrics. 2007 September ; 63(3): 901–909. doi:10.1111/j.1541-0420.2007.00762.x.

Multivariate Bernoulli Mixture Models with Application to Postmortem Tissue Studies in Schizophrenia

Zhuoxin Sun¹, Ori Rosen², and Allan R. Sampson³

¹Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, U.S.A

²Department of Mathematical Sciences, University of Texas at El Paso, El Paso, Texas 79968, U.S.A

³Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, U.S.A

Summary

A novel mixture model is presented for repeated measurements in which correlation among repeated observations on the same subject is induced via correlated unobservable component indicators. The mixture components in our model are linear regressions, and the mixing proportions are logits with random effects. Inference is facilitated by sampling from the posterior distribution of the parameters via Markov chain Monte Carlo methods. The model is applied to a neuronal postmortem brain tissue study to examine the differences in neuron volumes between schizophrenic and control subjects.

Keywords

MCMC; Mixture models; Multivariate Bernoulli distribution; Repeated measures; Schizophrenia

1. Introduction and Motivating Example

1.1 Overview

In this article, we introduce a novel model for repeated measures where each repeated observation has a mixture distribution. This model is motivated by our work with neuronal postmortem brain tissue studies, where a large number of neurons are sampled within a subject, and subject-level variables impact both the mixing proportions and the locations of the mixture components. Typically such a study considers more than 50 neurons per subject.

Our methodology is based on a multivariate extension of mixtures-of-experts, which is a mixture model for univariate variables proposed by Jacobs et al. (1991). In mixtures-of-

Correspondence to: Zhuoxin Sun.

Supplementary Materials

Web Appendices referenced in Sections 2, 3, and 4 are available under the Paper Information link at the Biometrics website <http://www.tibs.org/biometrics>.

experts, the mixture components are commonly generalized linear models while the mixing proportions are modeled as multivariate linear logits. Both the mixture components and the mixing proportions depend on covariates. Our multivariate model induces dependence by augmenting the data with component indicator variables within each subject that depend on both subject-specific random effects and experimental fixed effects.

Several extensions of mixture models have recently been proposed to account for dependence among repeated measurements. To model multiple eye-tracking observations from susceptible schizophrenic subjects, Rubin and Wu (1997) proposed a two-component mixture model in which the components are linear regressions with random effects, and the mixing proportions are linear logits. The within-subject dependence is accounted for by subject-specific random effects in the component distributions. We term this the Rubin–Wu model, although it is actually slightly less general than the “extra component mixture” model proposed in their paper for other purposes.

Other approaches to modeling dependent mixture response data include hidden Markov models (see McLachlan and Peel, 2000, Chapter 13) and mixtures of marginal models (Rosen, Jiang, and Tanner, 2000). The latter model combines the properties of mixtures-of-experts with those of generalized estimating equations (Liang and Zeger, 1986) and incorporates a working correlation matrix into each component to account for the dependence between observations on the same subject.

1.2 Motivating Example

A strong motivation for our model is a neuronal postmortem tissue study comparing schizophrenic and control subjects with regard to the somal volumes of deep layer 3 pyramidal neurons in the auditory association cortex (Sweet et al., 2003). In subjects with schizophrenia, the precision of the auditory sensory memory is usually deficient. Earlier studies indicate that imprecision of the auditory sensory memory may be related to abnormalities in the auditory association cortex. To further explore this result, Sweet et al. (2003) examined the somal volumes of deep layer 3 pyramidal cells in the auditory association cortex (Brodmann Area 42, BA42), using postmortem brain tissues from 18 schizophrenic subjects and 18 normal subjects. For each subject, three slide sections containing the region BA42 were selected by systematic random sampling. To sample cells on a slide section, random systematic sampling boxes were placed on the region of interest in each section, and the sampled cell volumes were obtained using the nucleator method (Gundersen, 1988). Approximately 100 to 250 neurons were selected in this manner for each subject. In layer 3 of BA42, some of the neurons have a longer axon and project to distant cortical regions. Other neurons have a shorter axon and project to the adjoining cortical region, Brodmann Area 41 (BA41). There is evidence that neuron volume is correlated with the extent of its axonal projection. This suggests that for each subject there might be within-region BA42 subgroups of neurons with different somal sizes. Sweet et al. (2003) treated all the observed neuron volumes as coming from one population and conducted a multivariate covariance analysis. These authors showed that the overall mean neuron volume decreases in schizophrenic subjects as compared to the control subjects. However, their approach was not able to detect a subgroup of neurons that are affected in subjects with schizophrenia. The

need to reexamine these data with a more biologically meaningful approach led us to consider a mixture model for somal volumes from BA42, where somal volumes measured within a subject are dependent. The neurobiologic causes of the dependence that is observed in these postmortem tissue studies are not well established. We have chosen to model the dependence by imposing a multivariate structure for the component indicator variables within a subject. It is known in this area of neuroscience that oftentimes a subject's age, gender, postmortem interval (PMI), and tissue storage time may affect neuron volumes and possibly the mixing proportions. Thus, in addition to the diagnosis main effect (schizophrenia or control), these additional covariates need to be taken into account for each subject.

In Section 2, we present our new model, multivariate Bernoulli mixtures of normals, and then compare it with normal component mixtures-of-experts and the Rubin–Wu model by examining the joint distribution of the observed data for each subject under each model. In Section 3, we develop a procedure for estimating the model parameters, using Markov chain Monte Carlo (MCMC) methods. In Section 4, we use our methodology to analyze the somal volume data. Simulation results are reported in Section 5, while possible extensions of our model and concluding remarks are given in Section 6.

2. Multivariate Bernoulli Mixtures of Normals

2.1 The Model

Let Y_{ij} and $\mathbf{x}_{ij}$ denote, respectively, the j th observation on subject i , and the covariate vector associated with observation Y_{ij} , where $i = 1, \dots, n$; $j = 1, \dots, I_i$. A latent component indicator variable for each observation Y_{ij} is denoted by Z_{ij} , which takes on values 0 and 1, corresponding to different populations. To describe the joint distribution of $(Z_{i1}, Z_{i2}, \dots, Z_{iI_i})^T$, let W_i , $i = 1, \dots, n$, be independent normally distributed random variables with mean 0 and variance σ_w^2 . Conditional on $W_i = w_i$, we assume that the Z_{ij} 's, $j = 1, \dots, I_i$, are independent Bernoulli random variables with mean

$$\pi(\mathbf{x}_{ij}, \boldsymbol{\gamma}, w_i) = \frac{e^{\mathbf{x}_{ij}^T \boldsymbol{\gamma} + w_i}}{1 + e^{\mathbf{x}_{ij}^T \boldsymbol{\gamma} + w_i}}. \quad (1)$$

Thus, we assume that these Bernoulli means are logits that depend on the covariate vector $\mathbf{x}_{ij}$ and random effect w_i . Marginally, the Z_{ij} 's are correlated, and $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iI_i})^T$ follows a multivariate Bernoulli distribution. The distribution of Y_{ij} , $i = 1, \dots, n$; $j = 1, \dots, I_i$, is given by

$$Y_{ij} | (Z_{ij}=1) \stackrel{\text{indep}}{\sim} N(\mathbf{x}_{ij}^T \boldsymbol{\beta}_1, \sigma_1^2)$$

$$Y_{ij} | (Z_{ij}=0) \stackrel{\text{indep}}{\sim} N(\mathbf{x}_{ij}^T \boldsymbol{\beta}_2, \sigma_2^2), \quad (2)$$

where $\beta_1, \beta_2, \gamma, \sigma_1^2, \sigma_2^2, \sigma_w^2$ are unknown parameters. The identifiability of the model is demonstrated in Web Appendix A.

2.2 The Joint Distribution of the Observed Data for Each Subject

To compare our model with the normal-component mixtures-of-experts and the Rubin–Wu model, we now derive a general form for the joint distribution of the observed data for each subject, encompassing all three models. For simplicity of notation and without loss of generality, we assume two observations on each subject, that is, $I_j = 2$.

The joint distribution of (Y_{j1}, Y_{j2}) for all three models can be expressed as

$$f_{y_{i1}, y_{i2}}(y_{i1}, y_{i2}) = \sum_{k=1}^2 \sum_{k'=1}^2 a_{kk'} \phi_2 \left((y_{i1}, y_{i2})^T; (\mu_{i1k}, \mu_{i2k'})^T, \Sigma_{kk'} \right), \quad (3)$$

where $\mu_{ijk} = x_{ij}^T \beta_k$, and $\mu_{ijk'} = x_{ij}^T \beta_{k'}$, for $j = 1, 2$, and $\phi_2(\cdot; \mu, \Sigma)$ is the bivariate normal density with mean vector μ and variance matrix Σ . For our model,

$a_{kk'} = \int \pi_{i1k} \pi_{i2k'} \phi(w_i; 0, \sigma_w^2) dw_i$, and

$$\Sigma_{kk'} = \begin{bmatrix} \sigma_k^2 & 0 \\ 0 & \sigma_{k'}^2 \end{bmatrix},$$

where $\pi_{ij1} = \pi_{ij}$, $\pi_{ij2} = 1 - \pi_{ij}$ and π_{ij} is as in equation (1), suppressing x_{ij} , γ , and w_i .

For mixtures-of-experts (Jacobs et al., 1991), the Z_{ij} 's are independent Bernoulli random variables with mean $p_{ij} = e^{x_{ij}^T \gamma} / (1 + e^{x_{ij}^T \gamma})$. For this model, $a_{kk'} = p_{i1k} p_{i2k'}$, and

$$\Sigma_{kk'} = \begin{bmatrix} \sigma_k^2 & 0 \\ 0 & \sigma_{k'}^2 \end{bmatrix},$$

where $p_{ij1} = p_{ij}$ and $p_{ij2} = 1 - p_{ij}$.

In the Rubin–Wu model, the Z_{ij} 's are the same as in mixtures-of-experts. Given Z_{ij} and a normally distributed subject-specific random effect S_i with mean 0 and variance σ_s^2 , where $S_1, \dots, S_n$ are independent, the conditional distribution of Y_{ij} for this model has form (3) with $a_{kk'} = p_{i1k} p_{i2k'}$, and

$$\Sigma_{kk'} = \begin{bmatrix} \sigma_k^2 + \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_{k'}^2 + \sigma_s^2 \end{bmatrix},$$

where $p_{ij1} = p_{ij}$ and $p_{ij2} = 1 - p_{ij}$.

Comparing the joint densities of the observed data under each model, we see that all three density functions are four-component mixtures of multivariate normals. In our model, the covariance matrices of the multivariate normal components are the same as those in the normal-component mixtures-of-experts; they have independent components with $\times_{j=1}^2 \{\sigma_1^2, \sigma_2^2\}$ describing all possible variances. In the Rubin–Wu model, the covariance matrix of each multivariate normal component has a compound-symmetric structure with σ_s^2 as the off-diagonal elements with variances described by $\times_{j=1}^2 \{\sigma_1^2 + \sigma_s^2, \sigma_2^2 + \sigma_s^2\}$. The mixing proportions of the joint densities in the normal-component mixtures-of-experts and those in the Rubin–Wu model correspond to independent random variables, whereas the mixing proportions in our model correspond to dependent multivariate Bernoulli random variables.

These results can be extended in a straightforward fashion to m observations on each subject. In this case, all the density functions can be written as 2^m mixtures of multivariate normals. Equation (3) provides insight into the structure of the data under our model, but does not usually aid in the estimation of the parameters. If we were to use classical maximum likelihood estimation based on the joint distribution of $Y_{i1}, \dots, Y_{im}$ directly, the difficulty of the problem would dramatically increase with the number of observations on each subject, making this computational approach infeasible for even moderate values of m . In our motivating example, m varies between 100 and 250, and for this reason we develop the Bayesian methodology described in the next section. We do not claim that a maximum likelihood approach to our model is impossible. In fact, maximum likelihood estimates could potentially be obtained via the EM or MCEM algorithm, using the augmented likelihood in Section 3.1, and integrating out numerically the w_i 's. However, recent references indicate that maximum likelihood estimates are not in general suitable for mixture modeling; for example, Gelman (2004) notes that “it is unstable to estimate mixture parameters by maximum likelihood.”

3. Inference

3.1 Augmented Likelihood and Prior Distributions

The hierarchical nature of our model lends itself naturally to Bayesian estimation via MCMC methods. We augment the observed data with the component indicators, Z_{ij} , $i = 1, \dots, n$, $j = 1, \dots, l_p$ and the subject-specific random effects, w_p , $i = 1, \dots, n$. Let $\mathbf{y} = (y_{11}, \dots, y_{1l_1}, \dots, y_{nl_1}, \dots, y_{nl_n})^T$, $\mathbf{z} = (z_{11}, \dots, z_{1l_1}, \dots, z_{n1}, \dots, z_{nl_n})^T$, and $\mathbf{w} = (w_1, \dots, w_n)^T$, so that the augmented likelihood is proportional to

$$\left(\sigma_w^2\right)^{-\frac{n}{2}} \prod_{i=1}^n \exp \left\{-\frac{w_i^2}{2\sigma_w^2}\right\} \times \prod_{j=1}^{l_i} \left[\pi_{ij} \frac{1}{\sqrt{\sigma_1^2}} \exp \left\{-\frac{(y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_1)^2}{2\sigma_1^2}\right\} \right]^{z_{ij}} \times \left[(1 - \pi_{ij}) \frac{1}{\sqrt{\sigma_2^2}} \exp \left\{-\frac{(y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_2)^2}{2\sigma_2^2}\right\} \right]^{1-z_{ij}}.$$

We place independent normal prior distributions on $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$ with means 0 and variance matrices $\sigma_\gamma^2 I_{q \times q}$, $\sigma_{\beta_1}^2 I_{q \times q}$, and $\sigma_{\beta_2}^2 I_{q \times q}$, respectively, where q is the length of the covariate vector. The priors on σ_1^2 , σ_2^2 and σ_w^2 are taken to be independent inverse gamma distributions, denoted by $\text{IG}(a_1, \delta_1)$, $\text{IG}(a_2, \delta_2)$, $\text{IG}(a_w, \delta_w)$, respectively, and the priors on $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\gamma}$

are assumed independent of those on σ_1^2 , σ_2^2 , and σ_w^2 . To obtain vague priors, the values of σ_γ^2 , $\sigma_{\beta_1}^2$, and $\sigma_{\beta_2}^2$ are assumed large, while α_1 , δ_1 , α_2 , δ_2 , α_w , and δ_w are set to small values. The specific values are given in Section 4.

3.2 The Sampling Scheme

We use MCMC methods to sample from the posterior distribution of the parameters. A Metropolis–Hastings step (Hastings, 1970) is performed for nonstandard conditional distributions. To achieve good mixing, we treat $\boldsymbol{\gamma}$ and $\boldsymbol{w}$ as a block, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ as another block. Sampling individually from the full conditional distributions of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$ and the random effects w_i 's, results in slow convergence, because of high correlation between $\boldsymbol{\gamma}$ and $\boldsymbol{w}$. The sampling scheme we use is as follows.

1. Initialize the parameters $\boldsymbol{\beta}_1^{(0)}$, $\boldsymbol{\beta}_2^{(0)}$, $\sigma_1^{2(0)}$, $\sigma_2^{2(0)}$, $\boldsymbol{\gamma}^{(0)}$, $\sigma_w^{2(0)}$ and $\boldsymbol{w}^{(0)}$. For iterations $t = 1, 2, \dots$:
2. Sample the component-indicators $z_{ij}^{(t+1)}$, $i = 1, \dots, n, j = 1, \dots, l_i$ from a Bernoulli random variable with mean $\tau_{ij}^{(t)}$, where

$$\tau_{ij}^{(t)} = \frac{\frac{1}{\sigma_1^{2(t)}} \exp \left\{ -\frac{(y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_1^{(t)})^2}{2\sigma_1^{2(t)}} + \mathbf{x}_{ij}^T \boldsymbol{\gamma}^{(t)} + w_i \right\}}{\frac{1}{\sigma_1^{2(t)}} \exp \left\{ -\frac{(y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_1^{(t)})^2}{2\sigma_1^{2(t)}} + \mathbf{x}_{ij}^T \boldsymbol{\gamma}^{(t)} + w_i \right\} + \frac{1}{\sigma_2^{2(t)}} \exp \left\{ -\frac{(y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_2^{(t)})^2}{2\sigma_2^{2(t)}} \right\}}. \quad (4)$$

3. Sample $\sigma_1^{2(t+1)}$ from $\text{IG}(\frac{1}{2} \sum_{i=1}^n l_i + \alpha_1, \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{l_i} z_{ij}^{(t+1)} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_1^{(t)})^2 + \delta_1)$.
4. Sample $\sigma_2^{2(t+1)}$ from $\text{IG}(\frac{1}{2} \sum_{i=1}^n l_i + \alpha_2, \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{l_i} z_{ij}^{(t+1)} (1 - z_{ij}^{(t+1)}) (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}_2^{(t)})^2 + \delta_2)$, such that $\sigma_2^{2(t+1)} < \sigma_1^{2(t+1)}$ for identifiability of the normal components.
5. Sample $\sigma_w^{2(t+1)}$ from $\text{IG}(\frac{n}{2} + \alpha_w, \frac{1}{2} \sum_{i=1}^n w_i^{2(t)} + \delta_w)$.
6. Sample $(\boldsymbol{\gamma}^{(t+1)}, \boldsymbol{w}^{(t+1)})$ as a block from their conditional distribution $p(\boldsymbol{\gamma}, \boldsymbol{w} | \mathbf{y}, \mathbf{z}^{(t+1)}, \boldsymbol{\beta}_1^{(t)}, \boldsymbol{\beta}_2^{(t)}, \sigma_1^{2(t+1)}, \sigma_2^{2(t+1)}, \sigma_w^{2(t+1)})$ via a Metropolis–Hastings step.
7. Sample $(\boldsymbol{\beta}_1^{(t+1)}, \boldsymbol{\beta}_2^{(t+1)})$ as a block from their conditional distribution $p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 | \mathbf{y}, \mathbf{z}^{(t+1)}, \sigma_1^{2(t+1)}, \sigma_2^{2(t+1)}, \boldsymbol{\gamma}^{(t+1)}, \boldsymbol{w}^{(t+1)}, \sigma_w^{2(t+1)})$ via a Metropolis–Hastings step.

Details of the Metropolis–Hastings steps are given in Web Appendix B.

When implementing this sampling scheme, some of the updated τ_{ij} given in equation (4) might be close to 0 or 1, implying that no observations, corresponding to some values of the covariates, are allocated to a mixture component. In this case, the Markov chain may move very slowly. To avoid this problem, we choose random starting values $\beta_1^{(0)}, \beta_2^{(0)}, \sigma_1^{2(0)}, \sigma_2^{2(0)}, \gamma^{(0)}, \sigma_w^{2(0)}$ and $w^{(0)}$ that guarantee that the initial τ_{ij} 's are away from 0 or 1, for all i and j . In the Metropolis–Hastings steps, the tuning constants of the proposal distributions, described in Web Appendix B, are selected such that the acceptance ratios for drawing the unknown parameters are at least 0.10. Some starting values may also result in low acceptance ratios, leading to slow convergence. For this reason, we examine the acceptance ratios in short preliminary runs to obtain the tuning constants and the appropriate starting values.

4. Application

In this section, we apply our model to the neuron volume data described in Section 1.2. Because somal volume distributions are typically right skewed, the standard neuroscience approach is to use the log transform of the neuronal volume and assume it follows a normal distribution (see, e.g., Tandrup, 2004; Marner, Soborg, and Pakkenberg, 2005). Histograms (and kernel density estimates) of the log-transformed neuron volumes for each of 36 subjects (18 schizophrenic subjects; 18 control subjects) are given in Figure 1. It can be seen from this figure that the data for most of the subjects appear to be bimodal or unimodal, thus visually supporting our mixture model. In addition, we fitted simple normal mixture models to 16 selected control subjects and tested for the number of components in each subject's mixture. Assuming that the transformed neuron volumes in each subject are independent and identically distributed, we used a bootstrapping method based on the likelihood ratio test statistic described in McLachlan and Peel (2000, Chapter 6) to test the number of components in each subject. For 13 subjects there was strong or moderate evidence supporting two components, and for three subjects there was no evidence for two or more components. It seems reasonable therefore to model a randomly chosen neuron as coming from one of two populations: a population of smaller neurons or a population of larger neurons, where these two populations are thought to, respectively, correspond to the projections to BA41 and BA42. We then treat the transformed somal volumes, y_{ij} , $i = 1, \dots, n$, $j = 1, \dots, I_i$ as coming from a mixture with two normal components and fit our model. Our results may not be robust to departures from normality, particularly if there is poor separation of the components. Each subject's covariate vector $\mathbf{x}_{ij}$ consists of an intercept, an indicator of diagnostic group (normal = 1, schizophrenic = 2), age, gender (female = 1, male = 2), PMI, and the corresponding tissue storage time. In this application, the covariates are not specific to the neuron, but to the subject.

To obtain vague prior distributions, we set $\sigma_\gamma^2 = \sigma_{\beta_1}^2 = \sigma_{\beta_2}^2 = 10$ in the normal priors on γ, β_1 , and β_2 , and take $\alpha_1 = \alpha_2 = \alpha_w = 0.01$ and $\delta_1 = \delta_2 = \delta_w = 0.02$ in the inverse gamma priors on σ_1^2, σ_2^2 , and σ_w^2 . The values used for the tuning constants in the Metropolis–Hastings steps are reported in Web Appendix B. Random initial values are selected such that $0.01 < \tau_{ij}^{(0)} <$

0.99, for all i and j , where $\tau_{ij}^{(0)}$ is the initial probability of $Z_{ij}=1$, given in equation (4). The MCMC algorithm is run for 148,000 cycles after a burn-in period of 2000 iterations, although the chain converges quickly and starts stabilizing after around 500 iterations. The algorithm is run three times starting from three different sets of random initial values. The results from all three runs agree very closely.

To avoid highly correlated values from the MCMC algorithm, we select one sample value from the chain every 100 values. Table 1 presents the posterior means and 95% credible intervals for the different parameters based on the resulting 1480 draws from the posterior distribution. Each parameter is estimated by its posterior mean obtained as the average of the corresponding 1480 draws. The 95% credible interval for each parameter is obtained by ordering the 1480 draws, and finding the 0.025 and 0.975 sample quantiles. The “smaller neuron population” and “larger neuron population” correspond to the two normal components in the model. The “mixing proportions” refer to the proportions of the smaller neurons.

The results of this analysis directly address the question in which Sweet et al. (2003) were interested. It is seen that for the smaller neuron population, the 95% credible interval for the diagnostic effect does not include zero, thereby indicating a significant diagnostic effect. The negative estimate indicates that subjects with schizophrenia have smaller volumes than controls for this population of neurons. For the larger neuron population there is no significant diagnostic effect, and for the proportion of smaller neurons (versus larger), there is no significant diagnostic difference. Our results suggest that the overall reduction for schizophrenic subjects in somal volume seen originally by Sweet et al. (2003) in the deep layer 3 pyramidal neurons (BA42) appears to be primarily due to a reduction in somal volume of this region’s smaller pyramidal neurons, a population presumably of locally projecting neurons. Obviously, to confirm further this result, other neurobiological studies need to be conducted.

Although of much less scientific interest, there were several other significant parameters. In the smaller neuron population, in addition to the strong diagnostic effect, there are significant gender and PMI effects, and storage time effect is marginally significant. In the larger neuron population, the age effect is significant, and the gender effect is marginally significant. Moreover, for the mixing proportions, gender has a significant effect, whereas storage time has a marginal effect. Additionally, we note that the significance of σ_w^2 in the mixing proportions confirms the dependence of the indicator components.

5. Simulation Study

In this section, we report the results of a simulation study conducted to ensure that the findings based on our methodology are valid, particularly for the neuron volume data set. We simulate 750 data sets, each of which has the “same” data structure as the neuron volume data. That is, in each data set, there are 36 subjects, each having the same number of repeated measures as in the neuron volume data. The response variable y_{ij} , $i = 1, \dots, 36$, $j = 1, \dots, I_j$ in each simulated data set can be viewed as a simulated log-transformed neuron

volume for the j th neuron in subject i . The corresponding covariate vector $\mathbf{x}_{ij}$ in the simulated data set is the same as the covariate vector in the neuron volume data.

Three sets of true parameter values for $\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, \gamma$, and σ_w^2 (see Table 2) are first chosen to represent the cases of well-, medium-, and poorly-separated multivariate Bernoulli mixture of normals. Based on McLachlan and Peel (2000, p. 9) and Schilling, Watkins, and Watkins (2002), the separation of two normal components can in general be assessed by $\Delta = |\mu_1 - \mu_2|/(\sigma_1 + \sigma_2)$ where the two components have means μ_1, μ_2 and variances σ_1^2, σ_2^2 , respectively. When $\sigma_2/\sigma_1 = 0.80$, Schilling et al. (2002) point out that the mixture density is bimodal if and only if $\Delta > 1.35, 1.26, 1.15, 1.01, 1.29$, respectively, corresponding to the mixing proportions $p = 0.3, 0.4, 0.5, 0.6, 0.7$. When $\sigma_2/\sigma_1 = 0.90$, the mixture density is bimodal if and only if $\Delta > 1.36, 1.25, 1.11, 1.16, 1.34$, respectively, corresponding to the mixing proportions $p = 0.3, 0.4, 0.5, 0.6, 0.7$. In our simulated data, each subject's neurons can be viewed as coming from a normal mixture where the components have means μ_{11}, μ_{12} and variances σ_1^2, σ_2^2 , respectively, where $\mu_{ik} = x_i^T \beta_k, k = 1, 2$, so that $\Delta_i = |\mu_{11} - \mu_{12}|/(\sigma_1 + \sigma_2)$. We assess separation in the simulated data by $\bar{\Delta} = (\sum_{i=1}^{36} n_i \Delta_i) / (\sum_{i=1}^{36} n_i)$, where n_i is the number of neurons for subject i . The $\bar{\Delta}$'s corresponding to the three sets of parameters are 4.1, 1.9, 1.1, respectively, and the σ_2/σ_1 ratios are 0.79, 0.89, and 0.89, respectively. Moreover, the corresponding mixing proportions for each subject have a wide range. Comparing these $\bar{\Delta}$'s with the minimum bimodal thresholds 1.01, 1.11, and 1.11, shows that the three sets of true values of $\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, \gamma$, and σ_w^2 correspond to well-, medium-, and poorly-separated mixtures of normals, respectively.

For each level of separation, our simulations result in 250 data sets. We fit our model to each simulated data set by using the sampling scheme described in Section 3.2. For a relatively small number of data sets, the overall acceptance ratios of (γ, w) were less than 0.10 due to poorly selected random starting values. This phenomenon tends to happen more in the medium and poorly separated cases, and these cases were not used in the simulations. For the remaining data sets, the posterior means and 95% credible intervals are obtained based on 8000 iterations after a 2000-iteration burn-in period. For most simulated data sets, the Markov chain stabilizes after 1000 iterations, while for some simulated data sets corresponding to poorly separated mixtures, the algorithm requires more burn-in cycles, for example, 2000 iterations, to achieve convergence.

For each of the three given sets of true values of $\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, \gamma$, and σ_w^2 , the average values and the mean square errors of each parameter across the 250 realizations are obtained, and the coverage rates of the 95% credible intervals for each parameter are provided in Table 2. This table shows that the estimates have little bias, and the coverage rates are reasonable. The mean square errors are overall small, except for those of the intercept in the mixing proportions and those of σ_w^2 . For the well-, medium-, and poorly-separated cases, Figures 2, 3, and 4, respectively, give histograms of the parameter estimates ($\hat{\beta}$'s) based on the 250 realizations for the diagnostic, age, gender, PMI, and storage time for each of the smaller and larger neuron populations, as well as the parameter estimates ($\hat{\gamma}$'s) in the mixing proportions. The vertical dash line is the true parameter value. Taken together, these

simulations suggest that the estimates and the inference obtained in Section 4 for the neuron volume data are valid. The validity of our modeling and estimating procedure in a complex setting as demonstrated in these simulations encouragingly suggests that our estimation approach would produce valid results for simpler data settings.

6. Discussion

The multivariate Bernoulli mixtures of normals described in this article can be generalized to multivariate Bernoulli mixture models where the mixture components are any member of the exponential family. The model fitting procedures given in Section 3 can be modified to fit, for example, multivariate Bernoulli mixtures of Poissons.

In this article, we have only considered two-component mixtures. Our results can also be extended to any finite number of components $g > 2$. In this case, in order to describe the component-indicator variables, we would construct families of multivariate multinomial distributions that depend on covariates. One approach to constructing such a distribution is to incorporate subject-specific random effects into the mixing proportions which are modeled by multivariate linear logits.

We are in the process of developing a generalized model where the mixture components are linear regressions with random effects while the mixing proportions are logistic regressions with another group of random effects. Thus, the within subject correlation is built into the model not only through the latent component indicator variables, but also through the component distributions. Very preliminary results applying this generalized model for the data set considered in Section 4 support the conclusions given in this section based on the current model.

In this article, the latent component indicator variables have been modeled by logit regression. Alternatively, these indicators could be modeled by probit regression as in Albert and Chib (1993). This would result in closed-form conditional distributions for $(\gamma, \mathbf{w})$. It is not clear, however, that this would result in better mixing in the algorithm.

In some cases where we do not have information about the number of components in the data, the number of components g may be treated as an unknown parameter and sampled from the posterior distribution using a reversible jump MCMC scheme, as proposed in Green (1995). This is a topic for further investigation.

In this article, we have fitted the multivariate Bernoulli mixtures of normals model to the neuron volume data to examine the diagnostic main effect. Although the neuron volume data contain between-subject factors only, our model can accommodate both between- and within-subject factors. Thus, our model can be applied to longitudinal, as well as to neurological studies.

Acknowledgments

The research of ZS was supported by NIMH grant 5P50 MH045156 15. The research of OR was supported by NIMH grant 5P50 MH045156 15 and NSF grant DMS04 05038. The research of ARS was supported by NIMH grant 5P50 MH045156 15 and NSF grant DMS00 72207. We wish to thank our collaborators, Drs Robert A. Sweet

and David A. Lewis (Department of Psychiatry, University of Pittsburgh) for their neurological insight and for allowing us to use the neuron volume data. We also wish to thank both referees, the associate editor, and the editor for their insightful comments and suggestions that greatly improved this article.

References

- Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*. 1993; 88:669–679.
- Gelman A. Parameterization and Bayesian modeling. *Journal of the American Statistical Association*. 2004; 99:537–545.
- Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*. 1995; 82:711–732.
- Gundersen HJ. The nucleator. *Journal of Microscopy*. 1988; 151:3–21. [PubMed: 3193456]
- Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 1970; 57:97–109.
- Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE. Adaptive mixtures of local experts. *Neural Computation*. 1991; 3:79–87.
- Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
- Marner L, Soborg C, Pakkenberg B. Increased volume of the pigmented neurons in the locus coeruleus of schizophrenic subjects: A stereological study. *Journal of Psychiatric Research*. 2005; 39:337C34. [PubMed: 15804384]
- McLachlan, GJ., Peel, D. *Finite Mixture Models*. New York: John Wiley & Sons; 2000.
- Rosen O, Jiang W, Tanner MA. Mixtures of marginal models. *Biometrika*. 2000; 87:391–404.
- Rubin DB, Wu Y. Modeling schizophrenic behavior using general mixture components. *Biometrics*. 1997; 53:243–261. [PubMed: 9147594]
- Schilling MF, Watkins AE, Watkins W. Is human height bimodal? *The American Statistician*. 2002; 56:223–229.
- Sweet RA, Pierri JN, Auh S, Sampson AR, Lewis DA. Reduced pyramidal cell somal volume in auditory association cortex of subjects with schizophrenia. *Neuropsychopharmacology*. 2003; 28:599–609. [PubMed: 12629543]
- Tandrup T. Unbiased estimates of number and size of rat dorsal root ganglion cells in studies of structure and cell survival. *Journal of Neurocytology*. 2004; 33:173–192. [PubMed: 15322376]

Figure 1.

Histograms (and kernel density estimates) of the log-transformed neuron volumes for each subject in the neuron volume data. The left two columns are histograms for control subjects; the right two columns are histograms for schizophrenic subjects.

Figure 2.

Well-separated mixtures of normals: Histograms of the parameter estimates ($\hat{\beta}$ s) from the 250 realizations for diagnostic, age, gender, PMI, and storage time in the smaller neuron population, larger neuron population, and the parameter estimates ($\hat{\gamma}$ s) in the mixing proportions. The vertical dash lines denote the true parameter values.

Figure 3. Medium-separated mixtures of normals: Histograms of the parameter estimates ($\hat{\beta}$ s) from the 250 realizations for diagnostic, age, gender, PMI, and storage time in the smaller neuron population, larger neuron population, and the parameter estimates ($\hat{\gamma}$ s) in the mixing proportions. The vertical dash lines denote the true parameter values.

Figure 4.

Poorly-separated mixtures of normals: Histograms of the parameter estimates ($\hat{\beta}$ s) from the 250 realizations for diagnostic, age, gender, PMI, and storage time in the smaller neuron population, larger neuron population, and the parameter estimates ($\hat{\gamma}$ s) in the mixing proportions. The vertical dash lines denote the true parameter values.

Table 1

Results of fitting the model to the neuron volume data. Estimates (posterior means) and 95% credible intervals. The algorithm was run for 148,000 iterations after a burn-in of 2000. Results are based on 1480 draws (selecting one iteration every 100).

	Posterior mean	2.5%	97.5%
Smaller neuron population			
Intercept (β_{10})	7.320	7.011	7.632
Diagnostic	-0.133	-0.194	-0.071
Age	0.002	-0.003	0.007
Gender	0.102	0.022	0.185
PMI	-0.017	-0.024	-0.008
Storage time	-0.000046	-0.000095	0.000004
σ_1^2	0.311	0.284	0.340
Larger neuron population			
Intercept (β_{20})	7.324	6.673	7.984
Diagnostic	-0.095	-0.230	0.043
Age	0.011	0.001	0.021
Gender	0.174	0.007	0.345
PMI	-0.001	-0.015	0.013
Storage time	0.000002	-0.000109	0.000114
σ_2^2	0.639	0.587	0.693
Mixing proportions			
Intercept (γ_0)	0.854	4.221	-2.485
Diagnostic	0.191	0.932	-0.511
Age	-0.008	0.046	-0.061
Gender	1.202	2.184	0.297
PMI	-0.065	0.022	-0.152
Storage time	-0.000626	-0.000025	-0.001277
σ_w^2	0.846	0.416	1.562

Table 2

Average estimates and coverage rates (Cover) based on 250 realizations for each of well-, medium-, and poorly-separated cases in the simulation study (mean square errors are in parentheses)

Parameters	Well separated			Medium separated			Poorly separated		
	True	Estimate	Cover	True	Estimate	Cover	True	Estimate	Cover
Smaller neurons									
Intercept (β_{10})	0.10	0.10 (0.22)	0.94	2.70	2.66 (0.17)	0.96	2.90	2.83 (0.29)	0.94
Diagnostic	0.30	0.30 (0.01)	0.95	0.01	0.00 (0.00)	0.96	0.08	0.08 (0.01)	0.93
Age	-0.50	-0.50 (0.00)	0.93	0.08	0.08 (0.00)	0.94	0.08	0.08 (0.00)	0.95
Gender	2.60	2.60 (0.01)	0.94	0.05	-0.06 (0.01)	0.93	-0.01	-0.02 (0.02)	0.95
PMI	0.80	0.80 (0.00)	0.93	0.02	0.02 (0.00)	0.94	0.04	0.04 (0.00)	0.94
Storage time	0.00	0.00 (0.00)	0.94	0.00	0.00 (0.00)	0.95	0.00	0.00 (0.00)	0.96
σ_1^2	8.00	7.99 (0.03)	0.94	2.50	2.50 (0.01)	0.93	2.50	2.52 (0.01)	0.94
Larger neurons									
Intercept (β_{20})	0.01	0.02 (0.22)	0.96	1.35	1.34 (0.07)	0.94	1.35	1.32 (0.12)	0.94
Diagnostic	1.20	1.20 (0.01)	0.94	0.12	0.12 (0.00)	0.95	0.12	0.12 (0.00)	0.94
Age	-0.10	-0.10 (0.00)	0.96	0.13	0.13 (0.00)	0.94	0.10	0.10 (0.00)	0.93
Gender	2.00	1.99 (0.01)	0.98	0.10	-0.10 (0.00)	0.96	-0.10	-0.11 (0.01)	0.92
PMI	0.90	0.90 (0.00)	0.98	0.13	0.13 (0.00)	0.94	0.11	0.11 (0.00)	0.96
Storage time	0.00	0.00 (0.00)	0.95	0.00	0.00 (0.00)	0.96	0.00	0.00 (0.00)	0.94
σ_2^2	5.00	5.01 (0.02)	0.95	2.00	2.01 (0.00)	0.93	2.00	2.00 (0.00)	0.96
Mixing prop.									
Intercept (γ_0)	-1.90	-0.92 (3.17)	0.98	1.90	0.56 (4.22)	0.99	1.90	0.78 (3.70)	0.99
Diagnostic	0.65	0.66 (0.23)	0.96	0.65	-0.62 (0.42)	0.97	-0.65	-0.60 (0.56)	0.96
Age	0.12	0.11 (0.00)	0.97	0.12	-0.11 (0.00)	0.99	-0.12	-0.12 (0.00)	0.97
Gender	-1.50	-1.40 (0.36)	0.93	1.50	1.33 (0.68)	0.96	1.50	1.46 (0.53)	0.98
PMI	-0.21	-0.23 (0.00)	0.93	0.21	0.23 (0.01)	0.94	0.21	0.25 (0.01)	0.95
Storage time	0.00	0.00 (0.00)	0.97	0.00	0.00 (0.00)	0.96	0.00	0.00 (0.00)	0.97

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Parameters	Well separated			Medium separated			Poorly separated		
	True	Estimate	Cover	True	Estimate	Cover	True	Estimate	Cover
σ_w^2	2.00	2.21 (0.40)	0.95	4.00	4.54 (2.15)	0.94	4.00	5.00 (3.44)	0.95