Technical Report RM669, Department of Statistics & Probability, Michigan State University

A BAYESIAN ANALYSIS OF HIERARCHICAL MIXTURES WITH APPLICATION TO CLUSTERING FINGERPRINTS*

By Sarat C. Dass and Mingfei Li

Michigan State University

Hierarchical mixture models arise naturally for clustering a heterogeneous population of objects where observations made on each object follow a standard mixture density. Hierarchical mixtures utilize complementary aspects of mixtures at different levels of the hierarchy. At the first (top) level, the mixture is used to perform clustering of the objects, while at the second level, nested mixture models are used as flexible representations of distributions of observables from each object. Inference for hierarchical mixtures is more challenging since the number of unknown mixture components arise in both the first and second levels of the hierarchy. In this paper, a Bayesian approach based on Reversible Jump Markov Chain Monte Carlo methodology is developed for the inference of all unknown parameters of hierarchical mixtures. Our methodology is then applied to the clustering of fingerprint images and used to assess the variability of quantities which are functions of the second level mixtures.

1. Hierarchical Mixture Models. Consider an object, $\mathcal{O}$, selected at random from a heterogenous population, $\mathcal{P}$, with G (unknown) clusters. Let $X \equiv (x_1, x_2, x_3, \cdots)$ denote the observables on $\mathcal{O}$ where $x_j \equiv (x_j^{(1)}, x_j^{(2)}, \cdots, x_j^{(d)})'$ is a d-variate random vector in R^d. A hierarchical mixture model for the distribution of $\mathcal{O}$ in the population is

$$q(\underline{x}) = \sum_{g=1}^{G} \omega_g \prod_{j=1}^{n} q_g(x_j) \tag{1.1}$$

where $\underline{x} = (x_1, x_2, \cdots, x_n)$ are the n observations made on $\mathcal{O}$, ω_g, $g = 1, 2, \cdots, G$ are the G cluster proportions with $\omega_g > 0$ and $\sum_{g=1}^{G} \omega_g = 1$, $q_g(\cdot)$ is the mixture density for the g-th cluster given by

$$q_g(x) = \sum_{k=1}^{K_g} p_{kg}\, f_{kg}(x \,|\, \theta_{kg}), \tag{1.2}$$

with f_{kg} denoting a density with respect to the Lebesgue measure on R^d, p_{kg} denoting the mixing probabilities satisfying: (1) $p_{kg} > 0$ and (2) $\sum_{k=1}^{K_g} p_{kg} = 1$, and θ_{kg} denoting the set of all unknown parameters in f_{kg}. Identifiability of the hierarchical mixture model of (1.1) is achieved by imposing the constraints

$$\omega_1 < \omega_2 < \cdots < \omega_G \quad \text{and} \quad \theta_{1g} \prec \theta_{2g} \prec \cdots \prec \theta_{K_g g}, \tag{1.3}$$

*Research supported under NSF DMS grant no. 0706385

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: Model-based clustering, Gaussian mixtures, Bayesian inference, Reversible Jump Markov Chain Monte Carlo methods, Fingerprint individuality

for each $g = 1, 2, \cdots, G$, where $\prec$ is a partial ordering to be defined later. The set of all unknown parameters in the hierarchical mixture model (1.1) is denoted by $\boldsymbol{x} = (G, \boldsymbol{\omega}, \boldsymbol{K}, \boldsymbol{p}, \boldsymbol{\theta})$ where $\boldsymbol{\omega} = (\omega_1, \omega_2, \cdots, \omega_G)$, $\boldsymbol{K} = (K_1, K_2, \cdots, K_g)$, $\boldsymbol{p} = (p_{kg}, k = 1, 2, \cdots, K_g,\ g = 1, 2, \cdots, G)$, and $\boldsymbol{\theta} = (\theta_{kg}, k = 1, 2, \cdots, K_g,\ g = 1, 2, \cdots, G)$.

Hierarchical mixture models of (1.1) arise naturally when the goal is to cluster a population of objects, where observables from each object follow a standard mixture distribution. At the first (top, or G) level, the mixture is used to perform clustering of the objects, while at the second (or K_g) level, nested mixture models (nested within each $g = 1, 2, \cdots, G$ specification) are used as flexible representations of distributions of observables. The unknown number of mixture components, or mixture complexity, arise at both levels of the hierarchy, and is, therefore, more challenging to estimate compared to standard mixtures.

Estimating mixture complexity has been the focus of intense research for many years resulting in various estimation methodologies in a broad application domain. Non-parametric methods were developed in Escobar and West (1995) and Roeder and Wasserman (1997), whereas Ishwaran et al. (2001) and Woo and Sriram (2006) developed methodology for the robust estimation of mixture complexity for count data. In a Bayesian framework, the work of Green and Richardson (1997) demonstrated that both issues of estimating mixture parameters can be unified if viewed as a problem of model selection. With this view in mind, Green and Richardson (1997) developed a Reversible Jump Markov Chain Monte Carlo (RJMCMC) approach for the estimating mixture complexity by exploring the space of models of varying dimensions.

In this paper, we develop a Bayesian framework based on RJMCMC for the inference on hierarchical mixture models. One advantage of our approach is that we are able to assess the variability of the cluster estimate, which in turn, can be used to ascertain the variability of quantities that are functions of the clusters. We present one such example based on clustering a sample of fingerprint images. One quantity that is of special interest is the probability of a random correspondence (PRC) which measures to what extent two randomly chosen fingerprints from a population match with each other. The RJMCMC methodology developed in this paper provides an estimate of its mean and variance.

In the subsequent text, we assume each f_{kg} is multivariate normal with mean vector $\boldsymbol{\mu}_{kg} \equiv (\mu_{kg}^{(1)}, \mu_{kg}^{(2)}, \cdots, \mu_{kg}^{(d)})' \in R^d$ and covariance matrix $\sum_{kg} \in R^d \times R^d$. Our analysis on fingerprint images in Section 6 revealed that it is adequate to consider diagonal covariance matrices of the form $\sum_{kg} = diag\left(\left(\sigma_{kg}^{(1)}\right)^2, \left(\sigma_{kg}^{(2)}\right)^2, \cdots, \left(\sigma_{kg}^{(d)}\right)^2\right)$ where $\left(\sigma_{kg}^{(b)}\right)^2$ is the variance of the b-th component of the multivariate normal distribution. Thus, we have

$$f_{kg}(x \,|\, \theta_{kg}) = \phi_d(x \,|\, \boldsymbol{\mu}_{kg}, \boldsymbol{\sigma}_{kg}) = \prod_{b=1}^{d} \phi_1\left(x^{(b)} \,|\, \mu_{kg}^{(b)}, \left(\sigma_{kg}^{(b)}\right)^2\right) \tag{1.4}$$

where $\phi_1(\cdot \,|\, \mu, \sigma^2)$ denotes the density of the univariate normal distribution with

mean μ and variance σ^2, and $\boldsymbol{\sigma}_{kg} \equiv \left(\left(\sigma_{kg}^{(1)}\right)^2, \left(\sigma_{kg}^{(2)}\right)^2, \cdots, \left(\sigma_{kg}^{(d)}\right)^2\right)'$ is the d-variate vector of the variances. The second identifiability condition of (1.3) is re-expressed in terms of the first component of the mean vector as

$$\mu_{1g}^{(1)} < \mu_{2g}^{(1)} < \cdots < \mu_{K_g g}^{(1)}. \tag{1.5}$$

In the subsequent text, the identifiability condition (1.5) based on the first components of $\boldsymbol{\mu}_{kg}$ for $k = 1, 2, \cdots, K_g$ will be re-written using the '$\prec$' symbol as

$$\boldsymbol{\mu}_{1g} \prec \boldsymbol{\mu}_{2g} \prec \cdots \prec \boldsymbol{\mu}_{K_g g} \tag{1.6}$$

for each $g = 1, 2, \cdots, G$.

For N independent objects selected randomly from the population, say, $\mathcal{O}_i \equiv (x_{ij},\, j = 1, 2, \cdots, n_i)$, $i = 1, 2, \cdots, N$, it follows that the joint distribution of the observations are

$$\prod_{i=1}^{N} q(\underline{x}_i) = \prod_{i=1}^{N} \sum_{g=1}^{G} \omega_g \prod_{j=1}^{n_i} \sum_{k=1}^{K_g} p_{kg}\, \phi_d\left(x_{ij} \,|\, \boldsymbol{\mu}_{kg}, \boldsymbol{\sigma}_{kg}\right). \tag{1.7}$$

Two other notations are introduced here: $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ will respectively denote the collection of all $\{\boldsymbol{\mu}_{kg},\, k = 1, 2, \cdots, K_g, g = 1, 2, \cdots, G\}$ and $\{\boldsymbol{\sigma}_{kg},\, k = 1, 2, \cdots, K_g, g = 1, 2, \cdots, G\}$ vectors. Our goal is to make inference about the unknown parameters $\boldsymbol{x} = (G, \boldsymbol{\omega}, \boldsymbol{K}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ based on the joint distribution in (1.7).

2. A Bayesian Framework for Inference. For the subsequent text, we introduce some additional notations. The symbol $I(\mathcal{S})$ denotes the indicator function of the set $\mathcal{S}$, that is $I(\mathcal{S}) = 1$ if $\mathcal{S}$ is true, and 0, otherwise. The notation $A, B, \cdots \,|\, C, D, \cdots$ denotes the distribution of random variables $A, B, \cdots$ conditioned on $C, D, \cdots$ with $\pi(A, B, \cdots \,|\, C, D, \cdots)$ denoting the specific form of the conditional distribution. The notation $\pi(A, B, \cdots \,|\, \cdots)$ denotes the distribution of $A, B, \cdots$ given the rest of the parameters. We specify a joint prior distribution on $\boldsymbol{x}$ as follows:

– (1) G and $\boldsymbol{K}$: We assume

$$\pi(G, \boldsymbol{K}) = \pi(G) \cdot \pi(\boldsymbol{K} \,|\, G) = \pi_0(G) \cdot \prod_{g=1}^{G} \pi_0(K_g) \tag{2.1}$$

where π_0 is the discrete uniform distribution between G_{min} and G_{max} (respectively, K_{min} to K_{max}), both inclusive, for G (respectively, K_g).

– (2) The first and second level mixing proportions: We assume

$$\pi(\boldsymbol{\omega},\, \boldsymbol{p} \,|\, G,\, \boldsymbol{K}) = G!\, D_G(\boldsymbol{\omega} \,|\, \delta_\omega) \cdot I(\omega_1 < \omega_2 < \cdots < \omega_G) \prod_{g=1}^{G} D_{K_g}(\boldsymbol{p}_g \,|\, \delta_p), \tag{2.2}$$

where $D_H(\cdot \,|\, \delta)$ denotes the H-dimensional Dirichlet density with the H-component baseline measure $(\delta, \delta, \cdots, \delta)$, where δ is a pre-specified constant, and $\boldsymbol{p}_g \equiv (p_{1g}, p_{2g}, \cdots, p_{Kg,g})'$.

The indicator function arises due to the imposed identifiability constraint (1.3) on $\boldsymbol{\omega}$; it follows that $G!$ is the appropriate normalizing constant for this constrained density which is obtained by integrating out $\boldsymbol{\omega}$ and noting that $D_G(\boldsymbol{\omega} \,|\, \delta_\omega)$ is invariant under different permutations of $\boldsymbol{\omega}$. $\omega_1, \omega_2, \cdots, \omega_G$ are exchangeable.
– (3) The prior on the mean vector is taken as

$$\begin{aligned}\pi(\boldsymbol{\mu} \,|\, \boldsymbol{K}, G) \;=\; & \prod_{g=1}^{G} \left[\left(K_g! \prod_{k=1}^{K_g} \phi_1(\mu_{kg}^{(1)} \,|\, \mu_0, \tau^2) \right) \cdot \left(I\left(\mu_{1g}^{(1)} < \mu_{2g}^{(1)} < \cdots < \mu_{K_g g}^{(1)} \right) \right) \right. \\ & \left. \times \left(\prod_{b=2}^{d} \prod_{k=1}^{K_g} \phi_1(\mu_{kg}^{(b)} \,|\, \mu_0, \tau^2) \right) \right]. \end{aligned} \tag{2.3}$$

The indicator function appears due to the identifiability constraint (1.3) imposed on $\boldsymbol{\mu}$ with resulting normalizing constant $K_g!$ for each $g = 1, 2, \cdots, G$.
– (4) The prior distribution of the variances is taken as

$$\pi(\boldsymbol{\sigma} \,|\, \boldsymbol{K}, G) = \prod_{g=1}^{G} \left(\prod_{k=1}^{K_g} \prod_{b=1}^{d} IG\left(\left(\sigma_{kg}^{(b)} \right)^2 \,|\, \alpha_0, \beta_0 \right) \right) \tag{2.4}$$

where IG denotes the inverse gamma distribution with prior shape and scale parameters α_0 and β_0, respectively. The joint prior distribution on $\boldsymbol{x} = (G, \boldsymbol{\omega}, \boldsymbol{K}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ is thus given by

$$\pi(\boldsymbol{x}) = \pi(G,\, \boldsymbol{K}) \cdot \pi(\boldsymbol{\omega},\, \boldsymbol{p} \,|\, G,\, \boldsymbol{K}) \cdot \pi(\boldsymbol{\mu} \,|\, G,\, \boldsymbol{K}) \cdot \pi(\boldsymbol{\sigma} \,|\, G,\, \boldsymbol{K}) \tag{2.5}$$

where the component priors are given by equations (2.1-2.4). The prior on $\boldsymbol{x}$ depends on the hyper-parameters δ_p,δ_ω,G_{max},G_{min},K_{min},K_{max}, μ_0 τ^2, α_0 and β_0, all of which need to be specified for a given application. This will be done in Sections 5 and 6.

The likelihood (1.7) involves several summations within each product term and is simplified by augmenting variables to denote the class labels of the individual observations. Two different class labels are introduced for the two levels of mixtures: The augmented variable $\boldsymbol{W} \equiv (W_1, W_2, \cdots, W_N)$ denotes the class label of the G sub-populations, that is, $W_i = g$ whenever object i, $\mathcal{O}_i$, arises from the g-th subpopulation, and $\boldsymbol{Z} \equiv (Z_1, Z_2, \cdots, Z_N)$ with $Z_i \equiv (Z_{ij}, j = 1, 2, \cdots, n_i)$, where $Z_{ij} = k$ for $1 \le k \le K_g$ if x_{ij} arises from the k-th mixture component $\phi_d(\cdot \,|\, \boldsymbol{\mu}_{kg}, \boldsymbol{\sigma}_{kg})$. We denote the augmented parameter space by the same symbol $\boldsymbol{x}$ as before, that is, $\boldsymbol{x} = (G, \boldsymbol{\omega}, \boldsymbol{K}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{W}, \boldsymbol{Z})$. The augmented likelihood is now

$$\ell(G, \boldsymbol{\omega}, \boldsymbol{K}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{W}, \boldsymbol{Z}) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} \prod_{g=1}^{G} \prod_{k=1}^{K_g} \left(\phi_d(x_{ij} \,|\, \boldsymbol{\mu}_{kg}, \boldsymbol{\sigma}_{kg}) \right)^{I(Z_{ij}=k, W_i=g)}. \tag{2.6}$$

with priors on $\boldsymbol{W}$ and $\boldsymbol{Z}$ given by

$$\pi(\boldsymbol{W}, \boldsymbol{Z} \,|\, G, \boldsymbol{K},\, \boldsymbol{\omega}, \boldsymbol{p}\,) = \pi(\boldsymbol{W} \,|\, G,\, \boldsymbol{\omega}) \cdot \pi(\boldsymbol{Z} \,|\, G,\, \boldsymbol{K},\, \boldsymbol{W}, \boldsymbol{p}\,) \tag{2.7}$$

where $\pi(\boldsymbol{W} \,|\, G, \boldsymbol{\omega}) = \prod_{i=1}^{N} \prod_{g=1}^{G} \omega_g^{I(W_i=g)}$ and $\pi(\boldsymbol{Z} \,|\, G,\, \boldsymbol{K},\, \boldsymbol{W}, \boldsymbol{p}\,) = \prod_{g=1}^{G} \prod_{i\,:\,W_i=g} \prod_{j=1}^{n_i} \prod_{k=1}^{K_g} p_{kg}^{I(Z_{ij}=k)}$. Based on the augmented likelihood and prior distributions, one can write down the posterior distribution (up to a normalizing constant) via Bayes Theorem. The posterior has the expression

$$\begin{aligned} \pi(\boldsymbol{x} \,|\, data) \quad &\propto \quad \ell(G, \boldsymbol{\omega}, \boldsymbol{K}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{W}, \boldsymbol{Z}) \times \pi(\boldsymbol{W}, \boldsymbol{Z} \,|\, G, \boldsymbol{K},\, \boldsymbol{\omega}, \boldsymbol{p}) \\ &\quad \times \pi(G, \boldsymbol{K}, \boldsymbol{\omega}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}) \end{aligned} \tag{2.8}$$

based on (2.5), (2.6) and (2.7).

3. Posterior Inference. The total number of unknown parameters in the hierarchical mixture model depends on the values G and $\boldsymbol{K}$. Thus, the posterior in (2.8) can be viewed as a probability distribution on the space of all hierarchical mixture models with varying dimensions. To obtain posterior inference for such a space of models, Green (1995) and Green and Richardson (1997) developed the RJMCMC for Bayesian inference. In this paper, we develop a RJMCMC approach to explore the posterior distribution in (2.8) resulting from the hierarchical mixture model specification. We briefly discuss the most general RJMCMC framework here. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be elements of the model space with possibly differing dimensions. The RJMCMC approach proposes a move, say m, with probability r_m. The move m takes $\boldsymbol{x}$ to $\boldsymbol{y}$ via the proposal distribution $q_m(\boldsymbol{x}, \boldsymbol{y})$. In order to maintain the time reversibility condition, we require to accept the proposal with probability

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min \left\{ 1, \frac{\pi(\boldsymbol{y} \,|\, data)}{\pi(\boldsymbol{x} \,|\, data)} \frac{r_{m'} q_{m'}(\boldsymbol{y}, \boldsymbol{x})}{r_m q_m(\boldsymbol{x}, \boldsymbol{y})} \right\}; \tag{3.1}$$

in (3.1), $q_{m'}(\boldsymbol{y}, \boldsymbol{x})$ represents the probability of moving from $\boldsymbol{y}$ to $\boldsymbol{x}$ based on the "reverse" move m', and $\pi(\boldsymbol{x} \,|\, data)$ denotes the posterior distribution of $\boldsymbol{x}$ given data. It is crucial that the moves m and m' be reversible (see Green (1995)), meaning that the densities $q_m(\boldsymbol{x}, \boldsymbol{y})$ and $q_{m'}(\boldsymbol{y}, \boldsymbol{x})$ have the same support with respect to a dominating measure. In case $\boldsymbol{y}$ represents the higher dimensional model, we can first sample $\boldsymbol{u}$ from a proposal $q_0(\boldsymbol{x}, \boldsymbol{u})$ (with possible dependence on $\boldsymbol{x}$), and then obtain $\boldsymbol{y}$ as a one-to-one function of $(\boldsymbol{x}, \boldsymbol{u})$. In that case, the proposal density $q_m(\boldsymbol{x}, \boldsymbol{y})$ in (3.1) is expressed as

$$q_m(\boldsymbol{x}, \boldsymbol{y}) = q_0(\boldsymbol{x}, \boldsymbol{u}) / \det \left[\frac{\partial \boldsymbol{y}}{\partial(\boldsymbol{x}, \boldsymbol{u})} \right] \tag{3.2}$$

where $\frac{\partial \boldsymbol{y}}{\partial(\boldsymbol{x}, \boldsymbol{u})}$ denotes the Jacobian of the transformation from $(\boldsymbol{x}, \boldsymbol{u})$ to $\boldsymbol{y}$, and det represents the absolute value of its determinant. If the triplet $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{y})$ involves some discrete components, then the Jacobian of the transformation is obtained by the one-to-one map of the continuous parts of $\boldsymbol{y}$ and $(\boldsymbol{x}, \boldsymbol{u})$, which can depend on the values realized by the discrete components.

For the inference on hierarchical mixture models, five types of updating steps are considered with reversible pairs of moves, (m, m'), corresponding to moves in

spaces of varying dimensions. The steps are:

$$\begin{cases} (1)\ \text{Update } G \text{ with } (m, m') \equiv (G\text{-split}, G\text{-merge}), \\ (2)\ \text{Update } \boldsymbol{K} \,|\, G, \boldsymbol{\omega}, \boldsymbol{W} \text{ with } (m, m') \equiv (K\text{-split}, K\text{-merge}), \\ (3)\ \text{Update } \boldsymbol{\omega} \,|\, G, \boldsymbol{K}, \boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \\ (4)\ \text{Update } \boldsymbol{W}, \boldsymbol{Z} \,|\, G, \boldsymbol{K}, \boldsymbol{\omega}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \text{ and} \\ (5)\ \text{Update } \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma} \,|\, G, \boldsymbol{K}, \boldsymbol{\omega}, \boldsymbol{W}, \boldsymbol{Z}. \end{cases} \tag{3.3}$$

The steps (3-5) do not involve jumps in spaces of varying dimensions, and can be carried out based on a regular Gibbs proposal.

3.1. *Update G.* To discuss the G-split and G-merge moves, we let $\boldsymbol{x}$ and $\boldsymbol{y}$ denote two different states of the model space, that is,

$$\boldsymbol{x} = (G, \boldsymbol{\omega}, \boldsymbol{K}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{W}, \boldsymbol{Z}) \quad \text{and} \quad \boldsymbol{y} = (G^*, \boldsymbol{\omega}^*, \boldsymbol{K}^*, \boldsymbol{p}^*, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^*, \boldsymbol{W}^*, \boldsymbol{Z}^*) \tag{3.4}$$

where the *s in (3.4) denote a possibly different setting of the parameters.

3.2. *The G-merge move.* The G-merge move changes the current G to $G-1$ (that is, $G^* = G-1$) and is carried out based on the following steps:

STEP 1: First, two of the G components, say g_1 and g_2 with $g_1 < g_2$, are selected randomly for merging into a new component g^*. The first level mixing proportions are merged as $\omega_{g^*} = \omega_{g_1} + \omega_{g_2}$.

STEP 2: The K-components in $\boldsymbol{K}$ corresponding to g_1 and g_2 are, respectively, K_{g_1} and K_{g_2}. These are combined to obtain the new K-value, K_{g^*}, in the following way. Adding $K_{g_1} + K_{g_2} = K_t$,

$$K_{g^*} = \begin{cases} (K_t+1)/2 & \text{if } K_t \text{ is odd, and} \\ K_t/2 & \text{if } K_t \text{ is even.} \end{cases} \tag{3.5}$$

STEP 3: Next, $(\boldsymbol{p}_{g_1}, \boldsymbol{\mu}_{g_1}, \boldsymbol{\sigma}_{g_1})$ and $(\boldsymbol{p}_{g_2}, \boldsymbol{\mu}_{g_2}, \boldsymbol{\sigma}_{g_2})$ are merged to obtain $(\boldsymbol{p}_{g^*}, \boldsymbol{\mu}_{g^*}, \boldsymbol{\sigma}_{g^*})$ as follows. The identifiability conditions of (1.6) holds for $g = g_1$ and $g = g_2$, and must be ensured to hold for $g = g^*$ after the merge step. To achieve this, the K_t $\boldsymbol{\mu}$'s are arranged in increasing order

$$\boldsymbol{\mu}_1 \prec \boldsymbol{\mu}_2 \prec \cdots \prec \boldsymbol{\mu}_{K_t-1} \prec \boldsymbol{\mu}_{K_t} \tag{3.6}$$

with associated probability p_j for $\boldsymbol{\mu}_j$, for $j = 1, 2, \cdots, K_t$. Thus, p_j are a re-arrangement of the K_t probabilities in $\boldsymbol{p}_{g_1}$ and $\boldsymbol{p}_{g_2}$ according to the partial ordering on $\boldsymbol{\mu}_{g_1}$ and $\boldsymbol{\mu}_{g_2}$ in (3.6). First, the case when K_t is even is considered. Adjacent $\boldsymbol{\mu}$ values in (3.6) are paired

$$\underbrace{\boldsymbol{\mu}_1 \prec \boldsymbol{\mu}_2} \prec \underbrace{\boldsymbol{\mu}_3 \prec \boldsymbol{\mu}_4} \prec \cdots \prec \underbrace{\boldsymbol{\mu}_{K_t-1} \prec \boldsymbol{\mu}_{K_t}}, \tag{3.7}$$

and the corresponding g^* parameters are obtained using the formulas $p^*_{kg^*} = \frac{p_{2k-1}+p_{2k}}{2}$,

$$\boldsymbol{\mu}^*_{kg^*} = \frac{p_{2k-1}\boldsymbol{\mu}_{2k-1} + p_{2k}\boldsymbol{\mu}_{2k}}{p_{2k-1}+p_{2k}}, \quad \text{and} \quad \boldsymbol{\sigma}^*_{kg^*} = \frac{p_{2k-1}\boldsymbol{\sigma}_{2k-1} + p_{2k}\boldsymbol{\sigma}_{2k}}{p_{2k-1}+p_{2k}}, \tag{3.8}$$

for $k = 1, 2, \cdots, K_{g^*}$.

STEP 4: To obtain $\boldsymbol{W}^*$ and $\boldsymbol{Z}^*$, objects with $W_i = g_1$ or $W_i = g_2$ are relabeled as $W_i^* = g^*$. For these objects, the allocation to the K_{g^*} components is carried out using a Bayes allocation scheme. The probability that object i is assigned to component k is

$$P(Z_{ij}^* = k \,|\, W_i^* = g^*) = \frac{p_{kg^*}^* \, \phi_d(x_{ij} \,|\, \boldsymbol{\mu}_{kg^*}^*, \boldsymbol{\sigma}_{kg^*}^*)}{\sum_{k=1}^{K_{g^*}} p_{kg^*}^* \, \phi_d(x_{ij} \,|\, \boldsymbol{\mu}_{kg^*}^*, \boldsymbol{\sigma}_{kg^*}^*)}. \tag{3.9}$$

for $k = 1, 2, \cdots, K_{g^*}$. The allocation of all x_{ij} to the K_{g^*} components is the product of the above probabilities, namely,

$$PmergeAlloc = \prod_{i \,:\, W_i^* = g^*} \prod_{j=1}^{n_i} P(Z_{ij}^* = k_{ij} \,|\, W_i^* = g^*) \tag{3.10}$$

where k_{ij} are the realized values of k when the allocation is done for each observation x_{ij}.

When K_t is odd, an index, i_0 is selected at random from the set of all odd integers up to K_t, namely, $\{1, 3, 5, \cdots, K_t\}$. The triplet $(p_{i_0}, \boldsymbol{\mu}_{i_0}, \boldsymbol{\sigma}_{i_0})$ is not merged with any other indices but the new $p_{i_0}^* = p_{i_0}/2$. The remaining adjacent indices are merged according to **STEP 3** above. For the G-merge step, the proposal density, $q_{m'}(\boldsymbol{x}, \boldsymbol{y})$, is given by

$$q_{m'}(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \binom{G}{2}^{-1} \times PmergeAlloc & \text{if } K_t \text{ is even, and} \\ \binom{G}{2}^{-1} \times PmergeAlloc \times \frac{2}{K_t + 1} & \text{if } K_t \text{ is odd.} \end{cases} \tag{3.11}$$

This completes the G-merge move.

3.3. *The G-split move.* The split move is reverse to the merge step above and is carried out in the following steps:

STEP 1: A candidate G-component for split, say g, is chosen randomly with probability $1/G$. The split components are denoted by g_1 and g_2. The first level mixing probability, ω_g, is split into ω_{g_1} and ω_{g_2} by generating a uniform random variable, u_0, in $[0, 1]$ and setting $\omega_{g_1} = u_0 \, \omega_g \quad \text{and} \quad \omega_{g_2} = (1 - u_0) \, \omega_g$.

STEP 2: The value of K_g is transformed to K_t where K_t is either $2K_g - 1$ or $2K_g$ with probability $1/2$ each. Once K_t is determined, a pair of indices (K_{g_1}, K_{g_2}) is selected randomly from the set of all possible pairs of integers in $\{K_{min}, K_{min} + 1, \cdots, K_{max}\}^2$ satisfying $K_{g_1} + K_{g_2} = K_t$. If M_0 is the total number of such pairs, then the probability of selecting one such pair is $1/M_0$. The selection of K_{g_1} and K_{g_2} determines the number of second level components in the g_1 and g_2 groups.

STEP 3: The aim now is to split each component of the triplet $(\boldsymbol{p}_g, \boldsymbol{\mu}_g, \boldsymbol{\sigma}_g)$ into 2 parts: $(\boldsymbol{p}_{g_1}, \boldsymbol{\mu}_{g_1}, \boldsymbol{\sigma}_{g_1})$ and $(\boldsymbol{p}_{g_2}, \boldsymbol{\mu}_{g_2}, \boldsymbol{\sigma}_{g_2})$ such that both $\boldsymbol{\mu}_{g_1}$ and $\boldsymbol{\mu}_{g_2}$ satisfy the constraints (1.6) for $g = g_1$ and g_2. The case of $K_{g_1} + K_{g_2} = 2K_g$ is first considered.

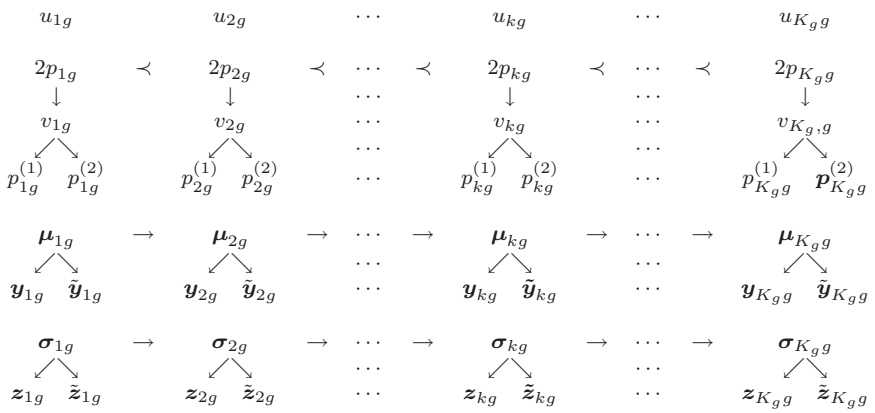


FIG 1. *Diagram showing the split of $2\boldsymbol{p}_g$, $\boldsymbol{\mu}_g$ and $\boldsymbol{\sigma}_g$. The partial ordering $\prec$ refers to the ordering of the $\mu_{kg}^{(1)}$s. The variables u_{kg}, $k = 1, 2, \cdots, K_g$ determine how many splits (out of two) go to component g_1 for each k. The right arrows '$\rightarrow$' represents the sequential split for $\boldsymbol{\mu}_g$ and $\boldsymbol{\sigma}_g$.*

A sketch of the split move is best described by the diagram in Figure 1, which introduce the additional variables to be used for performing the split. In Figure 1, $2\boldsymbol{p}_g$ is considered for splitting because the two split components will represent the second level mixing probabilities of g_1 and g_2, the sum of which together equals 2.

For each k, the variable u_{kg} in Figure 1 takes three values, namely, 0, 1 and 2 that respectively determines if the split components of $2p_{kg}$, $\boldsymbol{\mu}_{kg}$ and $\boldsymbol{\sigma}_{kg}$ either (1) both go to component g_2, (2) one goes to component g_1 and the other goes to g_2, or (3) both go to g_1. The variables u_{kg}, $k = 1, 2, \cdots, K_g$ must satisfy several constraints: (1) $\sum_{k=1}^{K_g} u_{kg} = K_{g_1}$, (2) $u_{kg} = 1$ for any k such that $p_{kg} > 0.5$, and (3) $\sum_{k\,:\,u_{kg}=h} 2p_{kg} < 1$ for $h = 0, 2$. Restriction (1) means that the number of components that go to g_1 must be K_{g_1} which is already pre-selected. The need for restriction (2) can be seen as follows: If $u_{kg} = 0$ or 2, and $p_{kg} > 0.5$, the total probability $2p_{kg}$ will be assigned to g_1 or g_2, and the sum of the second level mixing probabilities for that g component will be greater than 1, which is not possible. Restriction (3) is necessary to ensure that second level mixing probabilities for both g_1 and g_2 are non-negative (see equation (3.13)).

To generate the vector $\underline{u} \equiv (u_{1g}, u_{2g}, \cdots, u_{K_g g})'$, we consider all combinations of $\underline{u} \in \{0, 1, 2\}^{K_g}$, and reject the ones that do not satisfy the three restrictions. From the total number of remaining admissible combinations, M_1 say, we select a vector $\underline{u}$ randomly with equal probability $1/M_1$.

Once $\underline{u}$ has been generated, a random vector $\underline{v} \equiv (v_{kg},\, k = 1, 2, \cdots, K_g)$ is generated to split $2\boldsymbol{p}_g$ (see Figure 1). Some notations are in order: Let $A_0 = \{\, k : u_{kg} = 0\}$, $A_1 = \{\, k : u_{kg} = 1\}$ and $A_2 = \{\, k : u_{kg} = 2\}$. As in the case of $\underline{u}$, a few restrictions also need to be placed on the vector $\underline{v}$. To see what these restrictions are, we denote

$$p_{kg}^{(1)} = 2v_{kg}p_{kg} \quad \text{and} \quad p_{kg}^{(2)} = 2(1 - v_{kg})p_{kg} \tag{3.12}$$

for $k = 1, 2, \cdots, K_g$, to be the split components from $2\,p_{kg}$. Note that depending on the value of $u_{kg} = 0, 1$, or 2, the split components, $p_{kg}^{(1)}$ and $p_{kg}^{(2)}$, are either both assigned to component g_2, one to g_1 and the other to g_2, or both to g_1. For the case $u_{kg} = 1$, we will assume that $p_{kg}^{(1)}$ is the split probability that goes to g_1 and $p_{kg}^{(2)}$ goes to g_2. Note that the mixing probabilities for both components g_1 and g_2 should equal 1. This implies

$$\sum_{k\,:\,k\in A_1} p_{kg}^{(1)} + \sum_{k\,:\,k\in A_2} 2\,p_{kg} = 1 \quad \text{and} \quad \sum_{k\,:\,k\in A_1} p_{kg}^{(2)} + \sum_{k\,:\,k\in A_0} 2\,p_{kg} = 1 \tag{3.13}$$

for components g_1 and g_2, respectively. The second equation of (3.13) is redundant if the first is assumed since $\sum_{k\,:\,k\in A_1} p_{kg}^{(1)} + \sum_{k\,:\,k\in A_2} 2\,p_{kg} + \sum_{k\,:\,k\in A_1} p_{kg}^{(2)} + \sum_{k\,:\,k\in A_0} 2\,p_{kg} = 2\sum_{k=1}^{K_g} p_{kg} = 2$. We re-write the first equation as

$$\sum_{k\,:\,k\in A_1} a_k v_{kg} = 1 \tag{3.14}$$

where $a_k = 2p_{kg}/(1-\sum_{k\,:\,k\in A_2} 2p_{kg})$. Equation (3.14) implies that the entries of the vector $\underline{v}$ are required to satisfy two restrictions: (1) $0 \le v_{kg} \le 1$ for $k = 1, 2, \cdots, K_g$ from (3.12), and (2) Equation (3.14) above. In the Appendix, an algorithm is given to generate such a $\underline{v}$ where the proposal density can be written down in the closed form (see (7.2)).

The next step in the G-split move is to split $\boldsymbol{\mu}_g$ and $\boldsymbol{\sigma}_g$. Each component of $\boldsymbol{\mu}_g = (\boldsymbol{\mu}_{kg},\, k = 1, 2, \cdots, K_g)$ and $\boldsymbol{\sigma}_g = (\boldsymbol{\sigma}_{kg},\, k = 1, 2, \cdots, K_g$ in Figure 1 are split sequentially starting from $k = 1$, then $k = 2$ and so on until $k = K_g$. At the k-th stage, $\boldsymbol{\mu}_{kg}$ is split into the components $\boldsymbol{y}_{kg}$ and $\tilde{\boldsymbol{y}}_{kg}$ where $\boldsymbol{y}_{kg}$ is d-dimensional vector consisting of the entries $(y_{kg}^{(1)}, y_{kg}^{(2)}, \cdots, y_{kg}^{(d)})'$, and $\tilde{\boldsymbol{y}}_{kg} = (\tilde{y}_{kg}^{(1)}, \tilde{y}_{kg}^{(2)}, \cdots, \tilde{y}_{kg}^{(d)})'$. Similarly, $\boldsymbol{\sigma}_{kg}$ is split into the components $\boldsymbol{z}_{kg} \equiv (z_{kg}^{(1)}, z_{kg}^{(2)}, \cdots, z_{kg}^{(d)})'$ and $\tilde{\boldsymbol{z}}_{kg} \equiv (\tilde{z}_{kg}^{(1)}, \tilde{z}_{kg}^{(2)}, \cdots, \tilde{z}_{kg}^{(d)})'$. The collection of variables $\{\, \boldsymbol{y}_{kg},\, k = 1, 2, \cdots, K_g \,\}$ and $\{\, \boldsymbol{z}_{kg},\, k = 1, 2, \cdots, K_g \,\}$ are denoted by $\underline{y}$ and $\underline{z}$, respectively, and represent the additional variables that require to be generated for the split, via the proposal distribution $q_0(\underline{y}, \underline{z})$, say. The remaining variables (with ˜s) are obtained by solving the vector equations

$$\frac{p_{kg}^{(1)}\boldsymbol{y}_{kg} + p_{kg}^{(2)}\tilde{\boldsymbol{y}}_{kg}}{p_{kg}^{(1)} + p_{kg}^{(2)}} = \boldsymbol{\mu}_{kg} \quad \text{and} \quad \frac{p_{kg}^{(1)}\boldsymbol{z}_{kg} + p_{kg}^{(2)}\tilde{\boldsymbol{z}}_{kg}}{p_{kg}^{(1)} + p_{kg}^{(2)}} = \boldsymbol{\sigma}_{kg} \tag{3.15}$$

componentwise. We describe the split move further to see what properties $q_0(\underline{y}, \underline{z})$ should satisfy.

While the values of $\boldsymbol{y}_{kg}$ and $\tilde{\boldsymbol{y}}_{kg}$ (respectively, $\boldsymbol{z}_{kg}$ and $\tilde{\boldsymbol{z}}_{kg}$) are candidate values for $\boldsymbol{\mu}_{kg_1}$ and $\boldsymbol{\mu}_{kg_2}$ (respectively, $\boldsymbol{\sigma}_{kg_1}$ and $\boldsymbol{\sigma}_{kg_2}$), they are still not quite so since $\boldsymbol{\mu}_{g_1}$ and $\boldsymbol{\mu}_{g_2}$ must satisfy the constraints (1.6). To achieve this, we introduce two functions operating on d-dimensional vectors. The vector-valued "min" and "max" functions are defined as follows: For each $s = 1, 2, \cdots, S$, let $\boldsymbol{a}_s$ and

$\boldsymbol{b}_s$ denote two d-dimensional vectors given by $\boldsymbol{a}_s = (a_s^{(1)}, a_s^{(2)}, \cdots, a_s^{(d)})'$ and $\boldsymbol{b}_s = (b_s^{(1)}, b_s^{(2)}, \cdots, b_s^{(d)})'$. We define

$$\min(\boldsymbol{a}_s, \boldsymbol{b}_s) = \boldsymbol{a}_s \quad \text{and} \quad \max(\boldsymbol{a}_s, \boldsymbol{b}_s) = \boldsymbol{b}_s \tag{3.16}$$

for each $s = 1, 2, \cdots, S$ if $\boldsymbol{a}_1 \prec \boldsymbol{b}_1$ (recall this is by definition $a_1^{(1)} \leq b_1^{(1)}$), and vice versa when $\boldsymbol{b}_1 \prec \boldsymbol{a}_1$. Thus, the maximum and minimum functions above operate on the indices $s \geq 1$ with output depending on the index $s = 1$.

In the present case, consider the maximum and minimum functions defined as in (3.16) for each $k = 1, 2, \cdots, K_g$. Here, $S = 3$ with $\boldsymbol{a}_1 = \boldsymbol{y}_{kg}$, $\boldsymbol{a}_2 = \boldsymbol{z}_{kg}$, and $\boldsymbol{a}_3 = p_{kg}^{(1)}$, and $\boldsymbol{b}_1 = \tilde{\boldsymbol{y}}_{kg}$, $\boldsymbol{b}_2 = \tilde{\boldsymbol{z}}_{kg}$, and $\boldsymbol{b}_3 = p_{kg}^{(2)}$. If $\boldsymbol{y}_{kg} \prec \tilde{\boldsymbol{y}}_{kg}$, then it follows that

$$\begin{array}{rclcrcl} \min(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg}) & = & \boldsymbol{y}_{kg}, & & \max(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg}) & = & \tilde{\boldsymbol{y}}_{kg}, \\ \min(\boldsymbol{z}_{kg}, \tilde{\boldsymbol{z}}_{kg}) & = & \boldsymbol{z}_{kg}, & & \max(\boldsymbol{z}_{kg}, \tilde{\boldsymbol{z}}_{kg}) & = & \tilde{\boldsymbol{z}}_{kg}, \quad \text{and} \\ \min\left(p_{kg}^{(1)}, p_{kg}^{(2)}\right) & = & p_{kg}^{(1)}, & & \max\left(p_{kg}^{(1)}, p_{kg}^{(2)}\right) & = & p_{kg}^{(2)}; \end{array} \tag{3.17}$$

if $\tilde{\boldsymbol{y}}_{kg} \prec \boldsymbol{y}_{kg}$, then the opposite holds true.

To ensure that the constraints (1.6) hold, $\boldsymbol{y}_{kg}$ is generated in a way so that

$$\max\left(\boldsymbol{y}_{(k-1)g}, \tilde{\boldsymbol{y}}_{(k-1)g}\right) \prec \boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg} \prec \boldsymbol{\mu}_{(k+1)g} \tag{3.18}$$

in the sequential procedure for $k = 1, 2, \cdots, K_g$. In (3.18), maximum function $\max(\boldsymbol{y}_{0g}, \tilde{\boldsymbol{y}}_{0g})$ for $k = 1$ (respectively, $\boldsymbol{\mu}_{(K_g+1)g}$ for $k = K_g$) is defined to be the vector of lower (respectively, upper) bounds for the means. In the application to fingerprint images in Section 6, each image has size 500×500 which implies that the componentwise lower and upper bounds are, respectively, 0 and 500. For $\boldsymbol{z}_{kg}$ and $\tilde{\boldsymbol{z}}_{kg}$, we require these variables to satisfy the constraints $\boldsymbol{z}_{kg} \geq 0$ and $\tilde{\boldsymbol{z}}_{kg} \geq 0$ componentwise since they are candidate values for $\boldsymbol{\sigma}_{kg_1}$ and $\boldsymbol{\sigma}_{kg_2}$. Thus, the generation of $\underline{y}$ and $\underline{z}$, via the proposal distribution $q_0(\underline{y}, \underline{z})$, requires that (3.18), $\boldsymbol{z}_{kg} \geq 0$, and $\tilde{\boldsymbol{z}}_{kg} \geq 0$ be satisfied. A proposal density that achieves this is discussed in the Appendix.

The values of each triplet $(\boldsymbol{p}_{g_1}, \boldsymbol{\mu}_{g_1}, \boldsymbol{\sigma}_{g_1})$ and $(\boldsymbol{p}_{g_2}, \boldsymbol{\mu}_{g_2}, \boldsymbol{\sigma}_{g_2})$ can now be obtained. A sequential procedure is again adopted. The post-split parameters $\boldsymbol{p}_{g_h}$, $\boldsymbol{\mu}_{g_h}$ and $\boldsymbol{\sigma}_{g_h}$, $h = 1, 2$, are initialized to the empty set. Starting from $k = 1$, the sets are appended as follows: For $h = 0, 2$, define $h_1 = 2$ and $h_2 = 1$ if $h = 0$, and $h_1 = 1$ and $h_2 = 2$ if $h = 2$. For $k = 1, 2, \cdots, K_g$, if $k \in A_h$,

$$\begin{aligned} \boldsymbol{p}_{g_{h_1}} &= \left(\boldsymbol{p}_{g_{h_1}}, \min\left(p_{kg}^{(1)}, p_{kg}^{(2)}\right), \max\left(p_{kg}^{(1)}, p_{kg}^{(2)}\right)\right), \ \boldsymbol{p}_{g_{h_2}} = \left(\boldsymbol{p}_{g_{h_2}}\right), \\ \boldsymbol{\mu}_{g_{h_1}} &= \left(\boldsymbol{\mu}_{g_{h_1}}, \min\left(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg}\right), \max\left(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg}\right)\right), \ \boldsymbol{\mu}_{g_{h_2}} = \left(\boldsymbol{\mu}_{g_{h_2}}\right), \\ \boldsymbol{\sigma}_{g_{h_1}} &= \left(\boldsymbol{\sigma}_{g_{h_1}}, \min\left(\boldsymbol{z}_{kg}, \tilde{\boldsymbol{z}}_{kg}\right), \max\left(\boldsymbol{z}_{kg}, \tilde{\boldsymbol{z}}_{kg}\right)\right), \ \boldsymbol{\sigma}_{g_{h_2}} = \left(\boldsymbol{\sigma}_{g_{h_2}}\right), \end{aligned} \tag{3.19}$$

and if $k \in A_1$, $\boldsymbol{p}_{g_1} = \left(\boldsymbol{p}_{g_1}, p_{kg}^{(1)}\right)$, $\boldsymbol{p}_{g_2} = \left(\boldsymbol{p}_{g_2}, p_{kg}^{(2)}\right)$, $\boldsymbol{\mu}_{g_1} = (\boldsymbol{\mu}_{g_1}, \boldsymbol{y}_{kg})$, $\boldsymbol{\mu}_{g_2} = (\boldsymbol{\mu}_{g_2}, \tilde{\boldsymbol{y}}_{kg})$, $\boldsymbol{\sigma}_{g_1} = (\boldsymbol{\sigma}_{g_1}, \boldsymbol{z}_{kg})$, and $\boldsymbol{\sigma}_{g_2} = (\boldsymbol{\sigma}_{g_2}, \tilde{\boldsymbol{z}}_{kg})$.

The above procedure guarantees that the post-split components $\boldsymbol{\mu}_{g_1}$ and $\boldsymbol{\mu}_{g_2}$ satisfy the constraints (1.6). At this point, we can explicitly determine some of the components of $\boldsymbol{y}$ in equation (3.4); we have $G^* = G+1$, $\boldsymbol{K}^* = \boldsymbol{K} \cup \{K_{g_1}, K_{g_2}\} \setminus \{K_g\}$, $\boldsymbol{p}^* = \boldsymbol{p} \cup \{\boldsymbol{p}_{g_1}, \boldsymbol{p}_{g_2}\} \setminus \{\boldsymbol{p}_g\}$, $\boldsymbol{\mu}^* = \boldsymbol{\mu} \cup \{\boldsymbol{\mu}_{g_1}, \boldsymbol{\mu}_{g_2}\} \setminus \{\boldsymbol{\mu}_g\}$ and $\boldsymbol{\sigma}^* = \boldsymbol{\sigma} \cup \{\boldsymbol{\sigma}_{g_1}, \boldsymbol{\sigma}_{g_2}\} \setminus \{\boldsymbol{\sigma}_g\}$.

When $K_{g_1} + K_{g_2} = 2K_g - 1$, an index i_0 is selected from the set $\mathcal{I}_0 = \{\, k : 2p_{kg} < 1 \,\}$ with probability $1/|\mathcal{I}_0|$. The component with index i_0 is not split, and assigned a value of $u_{i_0 g}$ of either 0 or 2. For this case, $\underline{u} = (u_{1g}, u_{2g}, \cdots, u_{(K_g-1)g}, u_{i_0 g})$ is chosen from the product space $\{0,1,2\}^{(K_g-1)} \times \{0,2\}$, with M_1 denoting the number of admissible combinations satisfying the three restrictions on $\underline{u}$. After selecting a $\underline{u}$, we define $p^*_{i_0 g'} = 2p_{i_0 g'}$, $\boldsymbol{y}_{i_0 g'} = \tilde{\boldsymbol{y}}_{i_0 g'} = \boldsymbol{\mu}_{i_0 g'}$, and $\boldsymbol{z}_{i_0 g'} = \tilde{\boldsymbol{z}}_{i_0 g'} = \boldsymbol{\sigma}_{i_0 g'}$ where g' is either g_1 or g_2 depending on the selected $\underline{u}$. The split procedure above is carried out for the remaining indices $k \neq i_0$.

STEP 4: To complete the G-split proposal, we require to obtain the new first and second level labels, W^* and Z^*, in $\boldsymbol{y}$ (see (3.4)). All objects with labels $W_i = g$ are split into either $W_i^* = g_1$ or $W_i^* = g_2$ with allocation probabilities obtained as follows: Define $\mathcal{Q}_i(g_h) = \prod_{j=1}^{n_i} \sum_{k=1}^{K_{g_h}} p_{kg_h} \phi_d(x_{ij} \,|\, \boldsymbol{\mu}_{kg_h}, \boldsymbol{\sigma}_{kg_h})$ with $h = 1, 2$ for the i-th object. The W^*-allocation probabilities for components g_1 and g_2 are given by

$$P(W_i^* = g_h) = \mathcal{Q}_i(g_h)/(\mathcal{Q}_i(g_1) + \mathcal{Q}_i(g_2)) \tag{3.20}$$

for $h = 1, 2$. Once W_i^* has been determined, the Z_{ij}^*s are determined from the Bayes allocation probabilities (3.9) which is denoted here by $\mathcal{Q}_{ij}(k, g_h)$ for $h = 1, 2$. It follows that the allocation probability for the G-split move is

$$PsplitAlloc = \prod_{h=1,2} \prod_{i\,:\,W_i = g_{ih}} \mathcal{Q}_i(g_{ih}) \prod_{j=1}^{n_i} \mathcal{Q}_{ij}(k_{ij}, g_{ih}) \tag{3.21}$$

where g_{ih} is the realized value of g_h for the i-th object, and k_{ij} are the realized values of k for the second level labels Z_{ij}. Dass and Li (2008) give the proposal density for the G-split move as

$$q_m(\boldsymbol{x}, \boldsymbol{y}) = \frac{R_0\, q_0(\underline{v})\, q_0(\underline{y}, \underline{z}) \times PsplitAlloc}{G\, M_0\, M_1} \det\left[\frac{\partial \boldsymbol{y}}{\partial(\boldsymbol{x}, \boldsymbol{u})}\right] \tag{3.22}$$

where $R_0 = 1/2$ or $1/(2|\mathcal{I}_0|)$ according to whether $K_{g_1} + K_{g_2} = 2K_g$ or $2K_g - 1$ is chosen; in (3.22),

$$\det\left[\frac{\partial \boldsymbol{y}}{\partial(\boldsymbol{x}, \boldsymbol{u})}\right] = 2^{2(K_g-1)}\, \omega_g \left(\prod_{k \in A_0 \cup A_2 \cup A_2^c} p_{kg}\right) \prod_{k=1}^{K_g} \left(1 + \frac{p_{kg}^{(1)}}{p_{kg}^{(2)}}\right)^{2d} \tag{3.23}$$

is the absolute value of the Jacobian of the transformation from $(\boldsymbol{x}, \boldsymbol{u}) \to \boldsymbol{y}$, and A_1^c is the set A_1 excluding the largest element. The explicit expression of (3.23) is derived in the Appendix.

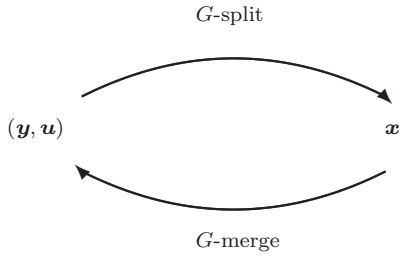


FIG 2. *Figure showing the G-split and G-merge proposals as a reversible pair of moves.*

We conclude the G-split and G-merge sections with a note on establishing reversibility of the two moves. The G-merge proposal (move m') takes $\boldsymbol{x}$ to $\boldsymbol{y}$ with proposal density given by $q_{m'}(\boldsymbol{x}, \boldsymbol{y})$ in (3.11). However, the acceptance probability in (3.1) also requires the proposal density, $q_m(\boldsymbol{y}, \boldsymbol{x})$, to move from $\boldsymbol{y}$ to $\boldsymbol{x}$ based on the "reverse" move m. In order to show that m is precisely the G-split move, we require to show that given $\boldsymbol{x}$ and $\boldsymbol{y}$, there is a unique $\boldsymbol{u}$ such that $\boldsymbol{x}$ can be obtained from the combination of $(\boldsymbol{y}, \boldsymbol{u})$ via the G-split move (see Figure 2). We demonstrate this in the next paragraph.

For the G-split move, the variables in $\boldsymbol{u}$ is given by $\boldsymbol{u} = (u_0, K_t, \underline{u}, \underline{v}, \underline{y}, \underline{z})$. These variables have the same interpretation as in the G-split move discussed earlier. Now, we check to see if $\boldsymbol{u}$ can be determined from $\boldsymbol{x}$ and $\boldsymbol{y}$. First, the variable u_0 can be determined from $u_0 = \omega_{g_1}/\omega_{g^*}$. Second, the value of $K_t = K_{g_1} + K_{g_2}$. Note that K_g alone cannot determine K_t since K_t is either $2K_g$ or $2K_g - 1$ with probability $1/2$ each. However, with information on K_{g_1} and K_{g_2}, K_t is uniquely determined.

Next, to get $\underline{u}$, we rearrange the components of $\boldsymbol{\mu}_{g_1}$ and $\boldsymbol{\mu}_{g_2}$ in the increasing order (3.6). If K_t is even, $K_g \equiv K_t/2$ adjacent pairs of $\boldsymbol{\mu}$s are formed, and u_{kg} is assigned the values $0, 1$ or 2 since it is known from which component (either g_1 or g_2) the two means in each of the k pairs came from. The case for odd K_t can be similarly handled since one of the $\boldsymbol{\mu}$ components is not paired, and subsequently, u_{kg} for that component is either 2 or 0 depending on whether the $\boldsymbol{\mu}$ component came from g_1 or g_2. Once $\underline{u}$ is obtained, $\underline{v}$ can be determined in the following way: The components of $\boldsymbol{p}_{g_1}$ and $\boldsymbol{p}_{g_2}$ are arranged according to the increasing order of $\boldsymbol{\mu}$s. Suppose the k-th pair consists of $\boldsymbol{\mu}_{\boldsymbol{k'g'}} \prec \boldsymbol{\mu}_{k''g''}$ with corresponding probabilities $p_{k'g'}$ and $p_{k''g''}$, where k', k'', g' and g'' are some indices of k and g resulting from the ordering. The value of $v_{kg} = p_{k'g'}/(2p_{kg})$ where $(p_{k'g'} + p_{k''g''})/2 = p_{kg}$ is the merged probability in the G-merge move.

Next, we obtain the values of $\boldsymbol{y}_{kg}$ and $\boldsymbol{z}_{kg}$. In the case $u_{kg} = 1$, $\boldsymbol{y}_{kg}$ (respectively, $\tilde{\boldsymbol{y}}_{kg}$) equals to the $\boldsymbol{\mu}$-value that came from component g_1 (respectively, g_2) in the pair $(\boldsymbol{\mu}_{k'g'}, \boldsymbol{\mu}_{k''g''})$. In the case when $u_{kg} = 0$ or 2, $\boldsymbol{y}_{kg}$ and $\tilde{\boldsymbol{y}}_{kg}$ can be determined only up to $\min(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg})$ and $\max(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg})$. Subsequently, the proposal density

$q_0(\underline{y}, \underline{z})$ depends only on $\min\left(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg}\right)$ and $\min\left(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg}\right)$ for these values of u_{kg}. A similar argument as above can be made for the $\boldsymbol{z}_{kg}$ and $\tilde{\boldsymbol{z}}_{kg}$.

3.4. *Update K.* We consider the move types K-split (type m) and K-merge (move type m'). The update of K is carried out for fixed G by selecting a component g on which K_g will be updated to either $K_g - 1$ or $K_g + 1$. For the K-merge move, we select two adjacent components for merging where adjacency is determined by the partial ordering (1.6). The merged mixing probability, mean and variance for the new component, k^*, are given by $p^*_{k^*g} = p_{kg} + p_{(k+1)g}$,

$$\boldsymbol{\mu}^*_{k^*g} = \frac{p_{kg}\boldsymbol{\mu}_{kg} + p_{k+1g}\boldsymbol{\mu}_{(k+1)g}}{p_{kg} + p_{(k+1)g}} \quad \text{and} \quad \boldsymbol{\sigma}^*_{k^*g} = \frac{p_{kg}\boldsymbol{\sigma}_{kg} + p_{k+1g}\boldsymbol{\sigma}_{(k+1)g}}{p_{kg} + p_{(k+1)g}}. \tag{3.24}$$

The objects with $W_i = g$ and $Z_{ij} = k$ or $(k+1)$ are merged into a newly relabelled bin $W^*_i = g$ and $Z^*_{ij} = k^*$.

For the K-split move, we first select a component that we want to split, k, which will be split into k_1 and k_2. A uniform random variable u_0 is selected to split p_{kg} into p_{k_1g} and p_{k_2g} in the following way:

$$p_{k_1g} = u_0 \cdot p_{kg} \qquad \text{and} \qquad p_{k_2g} = (1 - u_0) \cdot p_{kg}. \tag{3.25}$$

Next, $\boldsymbol{\mu}_{kg}$ and $\boldsymbol{\sigma}_{kg}$ are split by generating the variables $\boldsymbol{y}_{kg}$ and $\boldsymbol{z}_{kg}$ as in the case for the G-split move but now for a single k only. As in the G-split move, $\boldsymbol{y}_{kg}$ and $\tilde{\boldsymbol{y}}_{kg}$ are required to satisfy a similar constraint of the form

$$\boldsymbol{\mu}_{(k-1)g} \prec \boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg} \prec \boldsymbol{\mu}_{(k+1)g}, \tag{3.26}$$

so that the post-split $\boldsymbol{\mu}$ parameters satisfy the restriction (3.26), and subsequently (1.6). Once $\boldsymbol{y}_{kg}$ and $\boldsymbol{z}_{kg}$ are generated, the assignments to $\boldsymbol{\mu}_{k_hg}$ and $\boldsymbol{\sigma}_{k_hg}$, $h = 1, 2$ are done as follows:

$$\begin{aligned} \boldsymbol{\mu}_{k_1g} &= \min(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg}),\ \boldsymbol{\mu}_{k_2g} = \max(\boldsymbol{y}_{kg}, \tilde{\boldsymbol{y}}_{kg}), \\ \boldsymbol{\sigma}_{k_1g} &= \min(\boldsymbol{z}_{kg}, \tilde{\boldsymbol{z}}_{kg}),\ \boldsymbol{\sigma}_{k_2g} = \max(\boldsymbol{z}_{kg}, \tilde{\boldsymbol{z}}_{kg}). \end{aligned}$$

Objects with $W_i = g$ and observation labels $Z_{ij} = k$ are allocated to component k_1 or k_2 based on the Bayes allocation probabilities given by (3.9) with fixed $W_i = g$.

3.5. *Update Other Steps.* The update of the other quantities in steps (3-5) of equation (3.3) can be done via regular Gibbs sampler since they do not involve models in spaces of varying dimensions. We give the summary steps here for completion.

- Update $\boldsymbol{\omega} \,|\, G, \boldsymbol{K}, \boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}$: The conditional posterior distribution of $\boldsymbol{\omega}$ given the remaining parameters is given by

$$\pi(\boldsymbol{\omega} \,|\, \cdots) \propto \left(\prod_{g=1}^{G} \omega_g^{\delta_\omega + N_g - 1}\right) \cdot I(\omega_1 < \omega_2 < \cdots \omega_G) \tag{3.27}$$

where $N_g = \sum_{i=1}^{N} I(W_i = g)$ is the number of objects with label $W_i = g$. Equation (3.27) is the order statistic distribution of a Dirichlet with parameters $(\delta_\omega + N_1, \delta_\omega + N_2, \cdots, \delta_\omega + N_G)$ and can be easily simulated from.

• Update $\boldsymbol{W}, \boldsymbol{Z} \,|\, G, \boldsymbol{K}, \boldsymbol{\omega}, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}$: The conditional posterior distribution of $W_i, \boldsymbol{Z}_i$ is independent of each other. The update of W_i and $\boldsymbol{Z}_i \,|\, W_i$ based on the conditional posterior distribution is the Bayes allocation scheme of (3.20) and (3.9).

• Update $\boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma} \,|\, G, \boldsymbol{K}, \boldsymbol{\omega}, \boldsymbol{W}, \boldsymbol{Z}$: The conditional posterior distribution of $\boldsymbol{p} = \{\boldsymbol{p}_g,\ g = 1, 2, \cdots, G\}$ is given by

$$\pi(\boldsymbol{p} \,|\, \cdots) \propto \prod_{g=1}^{G} \prod_{k=1}^{K_g} p_{kg}^{(\delta_\pi + N_{kg} - 1)} \tag{3.28}$$

where $N_{kg} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} I(W_i = g, Z_{ij} = k)$ is the number of observations x_{ij} with $W_i = g$ and $Z_{ij} = k$. Thus, each $\boldsymbol{p}_g$ is independent Dirichlet with parameters $(\delta_p + N_{1g}, \delta_p + N_{1g}, \cdots, \delta_p + N_{K_g g})$. The update of $\boldsymbol{\mu}$ is carried out based on generating from its conditional posterior distribution $\pi(\boldsymbol{\mu} \,|\, \cdots)$. The generation scheme for $\boldsymbol{\mu}$ is as follows:

$$(\mu_{1g}^{(1)}, \mu_{2g}^{(1)}, \cdots, \mu_{K_g g}^{(1)}) \sim \left(\prod_{k=1}^{K_g} \phi_1(\mu_{kg}^{(1)} \,|\, \xi_{kg}^{(1)}, \eta_{kg}^{(1)}) \right) \cdot I\{\mu_{1g}^{(1)} < \mu_{2g}^{(1)} < \cdots < \mu_{K_g g}^{(1)}\} \tag{3.29}$$

independently for each $g = 1, 2, \cdots, G$, and for the remaining components,

$$\mu_{kg}^{(b)} \sim \phi_1(\mu_{kg}^{(b)} \,|\, \xi_{kg}^{(b)}, \eta_{kg}^{(b)}) \tag{3.30}$$

independently for each $b \geq 2$, $k = 1, 2, \cdots, K_g$ and $g = 1, 2, \cdots, G$; in (3.29) and (3.30),

$$\xi_{kg}^{(b)} = \frac{\frac{N_{kg}}{(\sigma_{kg}^{(b)})^2} \bar{x}_{kg}^{(b)} + \frac{1}{\tau^2}\mu_0}{\frac{N_{kg}}{(\sigma_{kg}^{(b)})^2} + \frac{1}{\tau^2}}, \quad \text{and} \quad \eta_{kg}^{(b)} = \left(\frac{N_{kg}}{\left(\sigma_{kg}^{(b)}\right)^2} + \frac{1}{\tau^2} \right)^{-1}. \tag{3.31}$$

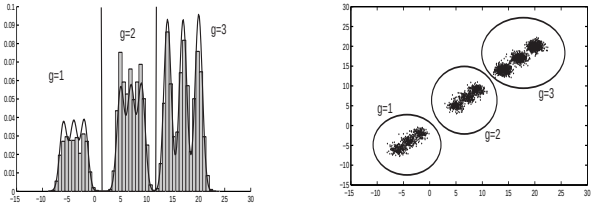
Equation (3.29) is the distribution of the order statistic from independent normals with different means and variances and can be simulated easily. The variances $\boldsymbol{\sigma}$, are updated via

$$\left(\sigma_{kg}^{(b)}\right)^2 \sim IG\left(\left(\sigma_{kg}^{(b)}\right)^2 \,|\, \alpha_{kg}^{(b)}, \beta_{kg}^{(b)} \right), \tag{3.32}$$

independently of each other, where

$$\alpha_{kg}^{(b)} = \alpha_0 + N_{kg} \quad \text{and} \quad \beta_{kg}^{(b)} = \left(1/\beta_0 + \sum_{ij} \left(x_{ij}^{(b)} - \mu_{kg}^{(b)} \right)^2 /2 \right)^{-1}$$

with $\sum_{ij}$ denoting the sum over all observations with $W_i = g$ and $Z_{ij} = k$.



(a) (b)

FIG 3. *(a) Density and (b) scatter plots for objects with univariate and bivariate observables corresponding to $d = 1$ and $d = 2$, respectively.*

3.6. *Update Empty Components.* The RJMCMC sampler developed also incorporates the updating of empty components into the chain. This is done with some modification to the earlier updating G and K move types. Empty components can arise naturally in the sampler when allocating the observations into the g or k components in both the G-split and K-split moves. In case of the G-split move, for example, it is possible that no objects are allocated into one of the split g components. Instead of rejecting this proposal altogether, we incorporate an additional variable, E_g, that indicates whether the g component is empty; $E_g = 1$ (respectively, 0) indicates that a component is non-empty (respectively, empty). The introduction of E_g incorporates additional steps into the RJMCMC algorithm, namely, E-Add and E-Remove which are reversible to each other. In the E-Remove move, an empty g^* component is selected for removal. The only change in this case is in the subpopulation parameters $\boldsymbol{\omega}$, since after the removal of ω_{g^*}, the remaining $\boldsymbol{\omega}$ probabilities should sum to 1. We thus have

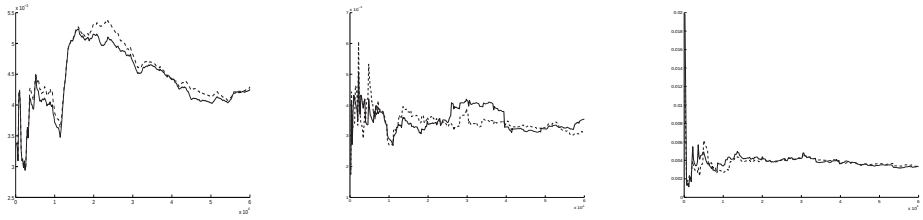
$$\omega_g^* = \omega_g/(1 - \omega_{g^*}) \tag{3.33}$$

for $g \neq g^*$. In the E-Add move type, a uniform random variable u_0 in $[0, 1]$ is generated and the probabilities $\boldsymbol{\omega}$ are redistributed to include the empty component g^* according to

$$\omega_{g^*} = u_0 \quad \text{and} \quad \omega_g^* = (1 - u_0) \cdot \omega_g \quad \text{for } g \neq g^*. \tag{3.34}$$

The proposal distributions for the E-Add and E-Remove move types and the associated Jacobians are given in the Appendix. The E-Add and E-Remove reversible move types for the K components is similar, and therefore, not discussed further.

4. Convergence Diagnostics. The assessment of convergence of the RJMCMC is carried out based on the methodology of Brooks and Guidici (1998,2000). Brooks and Guidici (1998,2000) suggests running $I \geq 2$ chains from different starting values and monitoring parameters that maintain the same interpretation across different models. Six quantities for used for monitoring, namely, the overall variance, $\hat{V}$, the within chain variance, W_c, within model variance W_m, within model within chain variance $W_m W_c$, the between model variance, B_m and the between model



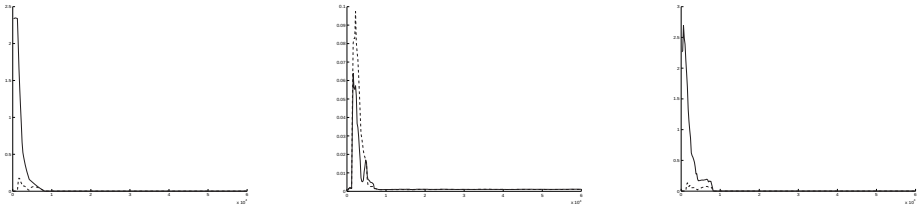
(a) (b) (c)

FIG 4. *Convergence diagnostics for $d = 1$. Panels (a), (b) and (c), respectively, show the plots of $(\hat{V}, W_c)$, $(W_m, W_m W_c)$ and $(B_m, B_m W_c)$ as a function of the iterations. The x-axis unit is 10,000 iterations.*

within chain variance, $B_m W_c$. For each monitoring parameter, the corresponding three plots of $\hat{V}$ and W_c, W_m and $W_m W_c$, and B_m and $B_m W_c$ against the number of iterations should be close to each other to indicate that the chains have sufficiently mixed. Our choice of the monitoring parameter is the log-likelihood of the hierarchical mixture model (see (1.7)).

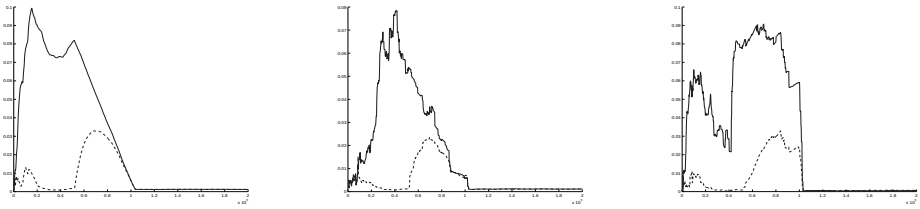
5. Simulation. Two simulation experiments were carried out for the cases $d = 1$ and $d = 2$ with prior parameter specifications given by $G_{min} = K_{min} = 2$, $G_{max} = K_{max} = 5$, $\delta_\pi = \delta_\omega = 1$, $\mu_0 = 7$, $\tau_0 = 20$, $\alpha_0 = 2.04$ and $\beta_0 = 0.5/(\alpha_0 - 1)$. The population of objects were simulated from $G = 3$ groups with population proportions $\boldsymbol{\omega} = (0.2, 0.3, 0.5)$. The nested K-components were chosen to be $K_g = 3$ for all $g = 1, 2, 3$. The specification of $\boldsymbol{p}$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are as follows: $\boldsymbol{p}_g = (0.33, 0.33, 0.34)$ for all g, $\boldsymbol{\mu}_1 = (-6, -4, -2)$, $\boldsymbol{\mu}_2 = (5, 7, 9)$ and $\boldsymbol{\mu}_3 = (14, 17, 20)$. Common variances were assumed: $\boldsymbol{\sigma}_g = (0.5, 0.5, 0.5)$ for $g = 1, 2, 3$. For the second experiment with $d = 2$, we took $\boldsymbol{\mu}_1 = \underline{\mathbf{1}} \cdot (-6, -4, -2)$, $\boldsymbol{\mu}_2 = \underline{\mathbf{1}} \cdot (5, 7, 9)$, and $\boldsymbol{\mu}_3 = \underline{\mathbf{1}} \cdot (14, 17, 20)$, where $\underline{\mathbf{1}} = (1, 1)'$. All component variances of $\boldsymbol{\sigma}$ were taken to be 0.5. The total number of objects sampled from the population were $N = 100$ with n_i (the number of observables from the i-th object) were iid from a Discrete Uniform distribution on the integers from 20 to 40, both inclusive. The density plot for the 3 components of the hierarchical mixture model in the case of $d = 1$ as well as the scatter plot for $d = 2$ based on a sample of observations from the population are given in Figures 3 (a) and (b), respectively. In both experimens, the RJMCMC algorithm is cycled through the 7 updating steps (5 steps in (3.3) as well as 2 steps involving updating empty G and K components). We took the probabilities of selecting various move types to be $r_m = r_{m'} = 0.5$ for the moves $(m, m') = (G\text{-split}, G\text{-merge})$ for $G = G_{min} + 1, G_{min} + 2, \cdots, G_{max} - 1$. When $G = G_{max}$, $r_m = 0 = 1 - r_{m'}$ and $r_{m'} = 0 = 1 - r_m$ for $G = G_{min}$. We used the same probabilities of selecting the K move types as above.

The space of hierarchical mixture models with the above specifications consists of $4^2 + 4^3 + 4^4 + 4^5 = 1,360$ models and so monitoring convergence of the chain becomes important. A total of $I = 3$ chains were chosen for monitoring with initial estimates of the hierarchical mixture model obtained using the values of $G = 2, 3$ and 4; Zhu,



(a) (b) (c)

FIG 5. *Convergence diagnostics for $d = 2$. Panels (a), (b) and (c), respectively, show the plots of $(\hat{V}, W_c)$, $(W_m, W_m W_c)$ and $(B_m, B_m W_c)$ as a function of the iterations. The x-axis unit is 10,000 iterations.*



(a) (b) (c)

FIG 6. *Convergence diagnostics for the NIST Fingerprint Database. Panels (a), (b) and (c), respectively, show the plots of $(\hat{V}, W_c)$, $(W_m, W_m W_c)$ and $(B_m, B_m W_c)$ as a function of the iterations. The x-axis unit is 20,000 iterations.*

Dass and Jain (2007) develops an algorithm that fits a hierarchical mixture model based on an agglomerative clustering procedure on the space of standard mixtures which requires a pre-specified value of G as input, and subsequently, the three choices of G mentioned were used to get the three initial estimates. The RJMCMC algorithm was run for $B = 60,000$ and convergence of the chain was monitored (see Figures 4 and 5). In both experiments, the RJMCMC appear to have converged after $60,000$ iterations. Highest posterior probabilities were found to be at the true values of the parameters.

6. An Application: Assessing the Individuality of Fingerprints. Fingerprint individuality refers to the study of the extent of uniqueness of fingerprints. It is the primary measure for assessing the uncertainty involved when individuals are identified based on their fingerprints, and has been the highlight of many court cases recently. In the case of Daubert v. Merrell Dow Pharmaceuticals (Daubert v. Merrell Dow Pharmaceuticals Inc., 1993), the U.S. Supreme Court ruled that in order for an expert forensic testimony to be allowed in courts, it had to be subject to five main criteria of scientific validation, that is, whether (i) the particular technique or methodology has been subject to statistical hypothesis testing, (ii) its error rates has been established, (iii) standards controlling the technique's operation exist and have been maintained, (iv) it has been peer reviewed, and (v)

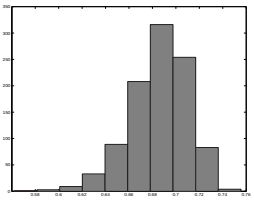


FIG 7. *Posterior distribution of* $\overline{PRC}$ *based on* $1,000$ *realizations of the RJMCMC after burn-in.*

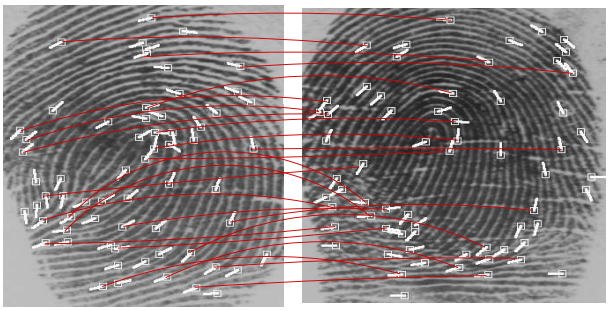


FIG 8. *Illustrating impostor minutiae matching (taken from Pankanti et al (2002)). A total of* $m = 64$ *and* $n = 65$ *minutiae were detected in left and right image, respectively, and* 25 *correspondences (i.e., matches) were found.*

it has a general widespread acceptance (see Pankanti, Prabhakar and Jain (2002), and Zhu et al (2007)). Following Daubert, forensic evidence based on fingerprints was first challenged in the 1999 case of U.S. v. Byron C. Mitchell, stating that the fundamental premise for asserting the uniqueness of fingerprints had not been objectively tested and its potential matching error rates were unknown. Subsequently, fingerprint based identification has been challenged in more than 20 court cases in the United States.

A quantitative measure of fingerprint individuality is given by the probability of a random correspondence (PRC), which is the probability that a random pair of fingerprints in the population will match with each other. Mathematically, the PRC is expressed as

$$PRC(\, w \,|\, m, n) = P(S \geq w \,|\, m, n\,), \tag{6.1}$$

where S denotes the number of feature matches with distribution based on all random pairs of fingerprints from the target population, w is the observed number of matches, and m and n, respectively, are the number of features in the two fingerprint images. Small (respectively, large) values of the PRC indicate low (respectively, high) uncertainty associated with the identification decision.

Here, we focus on a particular type of feature match based on minutiae. Minutiae are locations (i.e., $x_j \in R^2$) on the fingerprint image which correspond to

ridge anomalies (for example, ridge bifurcations and ridge endings) and are used by forensic experts to declare that two fingerprints belong to the same individual if sufficiently large number of minutiae are found to be common to both prints. Figure 8 shows an example of such a random match between two fingerprints of different individuals (also called an impostor match). The number of matches is determined by the number of minutiae in the right panel that falls within a square of area $4r_0^2$ centered at each minutiae in the left panel, where r_0 is a small number relative to the size of the fingerprint image. The number of matching minutiae in Figure 8 is 25, but it is not known how likely such a match occurs between a pair of impostor fingerprints in the population of individuals.

The reliability of the estimated PRC depends on how well elicited statistical models fits the distribution of minutiae in the population. Thus, candidate statistical models have to meet two important requirements: (i) flexibility, that is, the model can represent minutiae distributions over the entire population, and (ii) associated measures of fingerprint individuality can be easily obtained from these models. Zhu et al. (2007) demonstrated that a mixture of multivariate normals (with independent components) fits the distribution of minutiae well for each fingerprint. Furthermore, when m and n are large, the distribution of S in (6.1) can be approximated by a Poisson distribution with mean (expected) number of matches

$$\lambda(q_1, q_2, m, n) = m\, n\, p(q_1, q_2) \tag{6.2}$$

where q_h represents the normal mixture (see (1.2)) fitted to fingerprint h for $h = 1, 2$, and $p(q_1, q_2)$ is the probability of a match between a pair of random minutiae, one generated from q_1 and the other from q_2. The analytical expression for $p(q_1, q_2)$ is

$$p(q_1, q_2) \;=\; 4\, r_0^2 \sum_{k=1}^{K_1} \sum_{k'=1}^{K_2} \prod_{b=1}^{2} \phi_1 \left(0 \,|\, \underbrace{\left(\mu_{k1}^{(b)} - \mu_{k2}^{(b)}\right)}_{\mu}, \underbrace{\left(\sigma_{k1}^{(b)}\right)^2 + \left(\sigma_{k2}^{(b)}\right)^2}_{\sigma^2} \right), \tag{6.3}$$

where $\phi_1(\cdot \,|\, \mu, \sigma^2)$ is the normal density with mean μ and variance σ^2.

One drawback of Zhu et al (2007) is that no statistical model is elicited on the minutiae for a *population* of fingerprints. The hierarchical mixture model of (1.1) is such a population model on minutiae (since $x_j \in R^2$) satisfying both requirements of (i) flexibility and (ii) computational ease mentioned earlier. For a fingerprint pair coming from the subpopulation g_1 and g_2, we have $q_1 = q_{g_1}$ and $q_2 = q_{g_2}$ in (6.2). Hence, it follows that the mean PRC corresponding to w observed matches in the population is given by

$$\overline{PRC}(w \,|\, m, n\,) = \sum_{g_1=1}^{G} \sum_{g_2=1}^{G} \omega_{g_1}\, \omega_{g_2} P(S \geq w \,|\, \lambda(h_{g_1}, h_{g_2}, m, n)\,), \tag{6.4}$$

where S follows a Poisson distribution with mean $\lambda(h_{g_1}, h_{g_2}, m, n)$. The RJMCMC algorithm developed in the previous section can now be used to obtain the posterior distribution of $\overline{PRC}$. As an illustration, we considered 100 fingerprint images from the NIST Special Database 4 as a sample from a population of fingerprints. A total

of $I = 3$ chains were run with starting values given by the algorithm of Zhu et al. (2007) for the cases $G = 1, 2$ and 3. Figure 6 gives the diagnostics plots of the RJMCMC sampler which establish convergence after a burn-in of $B = 100,000$ runs. The posterior distribution of $\overline{PRC}$ (corresponding to $m = 64$, $n = 65$, $w = 25$ and $r_0 = 15$ pixels) based on $1,000$ realizations of the RJMCMC after the burn-in period is given in Figure 7. The posterior mean and standard deviation in Figure 7 is 0.6859 and 0.0250, respectively, and the 95% HPD interval is $[0.63, 0.735]$, approximately. We conclude that if a fingerprint pair was chosen from this population with $m = 64$, $n = 65$ and an observed number of matches $w = 25$, there is high uncertainty in making a positive identification. Our analysis actually indicates that the fingerprints in Figure 8 represent a typical impostor pair. The 95% HPD set suggests that the PRC can be as high as 0.735, that is, about 3 in every 4 impostor pairs can yield 25 or more matches.

7. Summary and Future Work. We have developed Bayesian inference methodology for the inference on hierarchical mixture models with application to the study of fingerprint individuality. Our future work will be to derive hierarchical mixture models on the extended feature space consisting of minutiae and other fingerprint features. In this paper, we only considered a two level hierarchy. The US-VISIT program now requires individuals to submit prints from all 10 fingers. This is the case of a 3-level hierarchical mixture model; in the first (top) level, individuals form clusters based on similar characteristics of their 10 fingers, and the distribution of features in each finger is modelled using standard mixtures. Hierarchical mixture models have potential use in other areas as well, including the clustering of soil samples (objects) based on soil characteristics (which can be modelled by a mixture or a transformation of mixtures).

Acknowledgment. The authors wish to acknowledge the support of NSF DMS grant no. 0706385 while conducting this research.

References.

[1] Brooks, S. P. and Giudici, P. (1998). Convergence assessment for Reversible Jump MCMC Simulations. *Bayesian Statistics 6*, Oxford University Press.

[2] Brooks, S. P. and Giudici, P. (2000). Markov Chain Monte Carlo Convergence Assessment vis Two-Way Analysis of Variance. *Journal of Computational and Graphical Statistics*, **9**, **2**, 266–285.

[3] Zhu, Y., Dass, S. C., and Jain, A. K. (2007). Statistical Models for Assessing the Individuality of Fingerprints. *IEEE Transactions on Information Forensics and Security*, **2**, **3**, 391–401.

[4] Escobar, M. and West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, **90**, **430**, 577–588.

[5] Roeder, K. and Wasserman, L. (1997). Practical Bayesian Density Estimation Using Mixtures of Normals. *Journal of the American Statistical Association*, **92**, **439**, 894–903.

[6] Woo, M. J. and Sriram, T. N (2007). Robust estimation of mixture complexity for count data. *Computational Statistics & Data Analysis*, **51**, **9**, 4379–4392.

[7] Ishwaran, H., James, L.F., and Sun, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *Journal of the American Statistical Association*, **96**, **456**, 1316-1332.

[8] Green, P. and Richardson, S. (1997). On the Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society Series B*, **59**, **4**, 731–792.

[9] Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, **4**, 711–732.
[10] NIST Special Database 4. NIST: 8-Bit Gray Scale Images of Fingerprint Image Groups (FIGS). *Online: http://www.nist.gov/srd/nistsd4.htm*
[11] Pankanti, S. and Prabhakar, S. and Jain, A. K. (2002). On the Individuality of Fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, **8**, 1010–1025.
[12] Daubert v. Merrel Dow Pharmaceuticals Inc. (1993). *509 U.S. 579, 113 S. Ct. 2786, 125 L.Ed.2d 469.*

Appendix.

7.1. *Generating* $\underline{v}$. The generation of $\underline{v}$ is discussed here. Let $|A_1| = T$ denote the cardinality of the set A_1 and let A_1^c be all k indices in A_1 excluding the largest one. Without loss of generality, we can relabel the k ordered indices in A_1 as $1, 2, \cdots, T$. It follows that the indices of A_1^c are $1, 2, \cdots, (T-1)$. The restriction $\sum_{k\,:\,k\in A_1} a_k v_k = 1$ can be rewritten as $v_T = (1 - \sum_{k=1}^{T-1} a_k v_k)/a_T$. Since $0 \le v_T \le 1$, it follows that the $(T-1)$ indices in A_1^c must satisfy the inequality $1 - a_T \le \sum_{k=1}^{T-1} a_k\, v_k \le 1$. Also note that $0 \le \sum_{k=1}^{T-1} a_k\, v_k \le \sum_{k=1}^{T-1} a_k$ since each $0 \le v_k \le 1$. Combining these inequalities, we get the following restriction on the $T-1$ free parameters of $\underline{v}$:

$$\max(0, 1 - a_T) \le \sum_{k=1}^{T-1} a_k v_k \le \min(1, \sum_{k=1}^{T-1} a_k) \tag{7.1}$$

Let $\mathcal{C} = \{\, (v_1, v_2, \cdots, v_{T-1}) \,:$ Equation (7.1) is satisfied and $0 \le v_k \le 1 \,\}$. It follows that $\mathcal{C}$ is a convex polyhedral in the unit hypercube $[0,1]^{T-1}$. The challenge now is to generate $(v_1, v_2, \cdots, v_{T-1})$ from $\mathcal{C}$ and be able to write down the proposal density $q_0(v_1, v_2, \cdots, v_{T-1})$ in a closed form. To do this, we determine the set of all extreme points of $\mathcal{C}$. There are a total of T inequality constraints on $(v_1, v_2, \cdots, v_{T-1})$: (1) $T-1$ constraints of the form $0 \le v_k \le 1$ and (2) one constraint of the form $A \le \sum_{k=1}^{T-1} a_k\, v_k \le B$ given by equation (7.1). Extreme points of a convex polyhedral in $T-1$ dimensions are formed by $T-1$ active equations. We can select $T-1$ candidate active equations from the T constraints above, solve for $(v_1, v_2, \cdots, v_{T-1})$ based on the $T-1$ equations and then check based on the remaining equation whether the solution obtained is admissible. For example, we may select all $v_k = 0$ for $k = 1, 2, \cdots, T-1$ from (1). Plugging in $(0, 0, \cdots, 0)$ in the remaining equation (2), we get $\sum_{k=1}^{T-1} a_k\, v_k = 0$. So, if $a_T < 1$ (respectively, $a_T \ge 1$), we get $0 < A$ (respectively, $0 \ge A$), giving an inadmissible (respectively, admissible) solution. Another candidate extreme point is formed by selecting the first $T-2$ v_k's to be zero and solving v_{T-1} from the equation $\sum_{k=1}^{T-1} a_k\, v_k = B$. The solution here is $(0, 0, 0, \cdots, 0, B/a_{T-1})$ and will be admissible if $0 \le B/a_{T-1} \le 1$ since the inequality that was not used is $0 \le v_{T-1} \le 1$.

It is easy to see that there are $(T-1)\, 2^{(T-1)}$ candidate extreme points to be checked for admissibility; we have to select $(T-1)$ constraints from T first which can be done in $T-1$ ways. Next, for each of the $T-1$ selected constraints, we can select either the lower or upper bounds of the constraints for candidate active equations.

The remaining constraint is used to check the admissibility of the candidate extreme point. Once this procedure is carried out, we obtain all admissible extreme points of $\mathcal{C}$, denoted by $\mathcal{E}$, say. From $\mathcal{E}$, we select the ones with smallest and largest v_1 values: v_{1L} and v_{1U}. The random variable v_1 is now generated from a uniform distribution between v_{1L} and v_{1U}. The uniform distribution represents the proposal distribution of v_1 and is denoted by $q_0(v_1)$.

Once v_1 is generated, equation (7.1) is updated to $\max(A - a_1\, v_1, 0) \le \sum_{k=2}^{T-1} a_k v_k \le \min(B - a_1\, v_1, \sum_{k=2}^{T-1} a_k)$ with $0 \le v_k \le 1$ for $k = 2, 3, \cdots, T-1$. Note that the new constraints again determine a convex polyhedral in $[0,1]^{(T-2)}$, and one can use the procedure outlined above to find all admissible extreme points. From these extreme points, we find the smallest and largest values of v_2, say $v_{2L}(v_1)$ and $v_{2U}(v_1)$ with proposal distribution $q_0(v_2 \,|\, v_1)$ which is uniform between $v_{2L}(v_1)$ and $v_{2U}(v_1)$. This procedure is carried out until $T-1$, and finally v_T is solved using equation $v_T = (1 - \sum_{k=1}^{T-1} a_k v_k)/a_T$. The proposal distribution for $\underline{v}$ is given by

$$q_0(\underline{v}) = \frac{1}{(v_{1U} - v_{1L})} \prod_{k=2}^{T-1} \frac{1}{(v_{kL}(v_1, v_2, \cdots, v_{k-1}) - v_{kU}(v_1, v_2, \cdots, v_{k-1}))} \tag{7.2}$$

Dass and Li (2008) also derives the proposal distributions for the other variables as well as the explicit form of the Jacobian $\det[\partial \boldsymbol{y}/\partial(\boldsymbol{x}, \boldsymbol{u})]$ for the G-split move. We refer the reader to Dass and Li (2008) for details.

7.2. *Generating $\underline{y}$ and $\underline{z}$.* Next, we consider the proposal densities for $\underline{y}$ and $\underline{z}$. The generation of $\boldsymbol{y}_{kg}$ is done sequentially starting from $k = 1, 2, \cdots, K_g$. At stage k, note that both $\boldsymbol{y}_{kg}$ and $\tilde{\boldsymbol{y}}_{kg}$ must satisfy the constraint (3.18), the lower bound of which we denote here by $Q_{(k-1)g} \equiv \max\left(\boldsymbol{y}_{(k-1)g}, \tilde{\boldsymbol{y}}_{(k-1)g}\right)$. From equation (3.18), we have

$$Q_{(k-1)g} \prec \boldsymbol{y}_{kg} \prec \boldsymbol{\mu}_{(k+1)g}$$

and

$$Q_{(k-1)g} \prec \frac{(p_{kg}^{(1)} + p_{kg}^{(2)})\boldsymbol{\mu}_{kg} - p_{kg}^{(1)} \boldsymbol{y}_{kg}}{p_{kg}^{(2)}} \prec \boldsymbol{\mu}_{(k+1)g}.$$

Solving the above two inequalities for $\boldsymbol{y}_{kg}$, we get the following upper and lower bounds for $\boldsymbol{y}_{kg}$:

$$A^* \prec \boldsymbol{y}_{kg} \prec B^*$$

where

$$A^* = \max\left(\boldsymbol{\mu}_{kg} - \frac{p_{kg}^{(2)}}{p_{kg}^{(1)}}(\boldsymbol{\mu}_{(k+1)g} - \boldsymbol{\mu}_{kg}), Q_{(k-1)g}\right)$$

and

$$B^* = \min\left(\boldsymbol{\mu}_{kg} + \frac{p_{kg}^{(2)}}{p_{kg}^{(1)}}(\boldsymbol{\mu}_{kg} - Q_{(k-1)g}), \boldsymbol{\mu}_{(k+1)g}\right).$$

Since $\prec$ puts restriction on the first component of $\boldsymbol{y}_{kg}$, $y_{kg}^{(1)}$ is generated from a uniform distribution between the first components of A^* and B^*, $A^{(1)^*}$ and $B^{(1)^*}$ say, when $u_{kg} = 1$. As seen before, when $u_{kg} = 0$ or 2, $y_{kg}^{(1)}$ and $y_{kg}^{(2)}$ are known only up to $\min(y_{kg}^{(1)}, y_{kg}^{(2)})$ and $\max(y_{kg}^{(1)}, y_{kg}^{(2)})$. In this case, $\min(y_{kg}^{(1)}, y_{kg}^{(2)})$ is generated from the uniform distribution between $A^{(1)^*}$ and $(A^{(1)^*} + B^{(1)^*})/2$. The remaining $y_{kg}^{(b)}$ are generated as iid uniform between the lower and upper bounds for the means.

The case of $\boldsymbol{z}_{kg}$ is similar. Both $\boldsymbol{z}_{kg}$ and $\tilde{\boldsymbol{z}}_{kg}$ must satisfy $\boldsymbol{z}_{kg} \geq 0$ and $\tilde{\boldsymbol{z}}_{kg} \geq 0$ componentwise. Based on a similar analysis above, it follows that each component $\boldsymbol{z}_{kg}^{(b)}$ must satisfy

$$0 \leq z_{kg}^{(b)} \leq \left(\sigma_{kg}^{(b)}\right)^2 \left(1 + \frac{p_{kg}^{(2)}}{p_{kg}^{(1)}}\right) \tag{7.3}$$

Thus, we propose to generate $\underline{z}$ from independent $IG(\cdot \mid \alpha_0, \beta_0)$ distributions subject to (7.3). It follows that the proposal density for $\underline{z}$ is

$$q_0(\underline{z}) \sim \prod_{k=1}^{K_g} \prod_{b=1}^{B} IG(z_{kg}^{(b)} \mid \alpha_0, \beta_0) \cdot I\left(0 \leq z_{kg}^{(b)} \leq \left(\sigma_{kg}^{(b)}\right)^2 \left(1 + \frac{p_{kg}^{(2)}}{p_{kg}^{(1)}}\right)\right). \tag{7.4}$$

7.3. *The Jacobian $\partial \boldsymbol{y}/\partial(\boldsymbol{x}, \boldsymbol{u})$.* For the G-split move with g being the component that was selected for split, we can focus only on the components of $\boldsymbol{y}$, $\boldsymbol{x}$ and $\boldsymbol{u}$ that were changed. So, we have

$$\boldsymbol{x} = (G, \omega_g, K_g, \boldsymbol{p}_g, \boldsymbol{\mu}_g, \boldsymbol{\sigma}_g), \tag{7.5}$$

$$\boldsymbol{u} = (u_0, K_t, \underline{u}, \underline{v}, \boldsymbol{y}_g, \boldsymbol{z}_g) \tag{7.6}$$

and

$$\boldsymbol{y} = (G+1, \omega_{g_1}, \omega_{g_2}, K_{g_1}, K_{g_2}, \boldsymbol{p}_{g_1}, \boldsymbol{p}_{g_2}, \boldsymbol{\mu}_{g_1}, \boldsymbol{\mu}_{g_2}, \boldsymbol{\sigma}_{g_1}, \boldsymbol{\sigma}_{g_2}). \tag{7.7}$$

The Jacobian matrix (conditioned on $\underline{u}$ and K_t) can be written in the following form:

$\boldsymbol{y}/(\boldsymbol{x}, \boldsymbol{u})$	ω_g	u_0	$\boldsymbol{p}_g$	$\underline{v}$	$\boldsymbol{\mu}_g$	$\underline{y}$	$\boldsymbol{\sigma}_g$	$\underline{z}$
ω_{g_1}	$\star$	$\star$	0	0	0	0	0	0
ω_{g_2}	$\star$	$\star$	0	0	0	0	0	0
$\boldsymbol{p}_{g_1}$	0	0	$\star$	$\star$	0	0	0	0
$\boldsymbol{p}_{g_2}$	0	0	$\star$	$\star$	0	0	0	0
$\boldsymbol{\mu}_{g_1}$	0	0	$\star$	$\star$	$\star$	$\star$	0	0
$\boldsymbol{\mu}_{g_2}$	0	0	$\star$	$\star$	$\star$	$\star$	0	0
$\boldsymbol{\sigma}_{g_1}$	0	0	$\star$	$\star$	0	0	$\star$	$\star$
$\boldsymbol{\sigma}_{g_2}$	0	0	$\star$	$\star$	0	0	$\star$	$\star$

(7.8)

where $\star$ indicate non-zero entries. It is clear that the Jacobian is a block lower triangular matrix, and so, its determinant depends on the determinant of the diagonal blocks. The diagonal determinants of the first two blocks are of

$$\begin{aligned} \omega_{g_1} &= u_0 \omega_g \\ \omega_{g_2} &= (1 - u_0)\omega_g \end{aligned} \tag{7.9}$$

with $\det\left[\frac{\partial(\omega_{g_1},\omega_{g_2})}{\partial(\omega_g,u_0)}\right]=\omega_g$, and

$$\begin{aligned} p_{kg}^{(1)} &= 2v_{kg}p_{kg} \\ p_{kg}^{(2)} &= 2(1-v_{kg})p_{kg} \end{aligned} \tag{7.10}$$

for $k=1,2,\cdots,K_g$ with $\det\left[\frac{\partial(p_{kg}^{(1)},p_{kg}^{(2)})}{\partial(p_{kg},v_{kg})}\right]=2^{2(K_g-1)}\prod_{k\in A_0\cup A_2\cup A_2^c}p_{kg}$. Note that in (7.10), the last equation with $k=K_g$ is redundant due to the restrictions placed on $\underline{v}$. There are a total of $|A_0|+|A_2|+|A_1|-1$ free parameters, and that is why the Jacobian involves K_g-1 and $A_0\cup A_2\cup A_1^c$, and not K_g and $A_0\cup A_2\cup A_1$, respectively. The determinant for the block of μ is given by

$$\left.\begin{aligned} \boldsymbol{\mu}_{k'g_1} &= \min(\boldsymbol{y}_{kg},\tilde{\boldsymbol{y}}_{kg}) \\ \boldsymbol{\mu}_{(k'+1)g_1} &= \max(\boldsymbol{y}_{kg},\tilde{\boldsymbol{y}}_{kg}) \end{aligned}\right\} \quad \text{if } u_{kg}=2, \tag{7.11}$$

$$\left.\begin{aligned} \boldsymbol{\mu}_{k'g_2} &= \min(\boldsymbol{y}_{kg},\tilde{\boldsymbol{y}}_{kg}) \\ \boldsymbol{\mu}_{(k'+1)g_2} &= \max(\boldsymbol{y}_{kg},\tilde{\boldsymbol{y}}_{kg}) \end{aligned}\right\} \quad \text{if } u_{kg}=0, \tag{7.12}$$

and if $u_{kg}=1$,

$$\begin{aligned} \boldsymbol{\mu}_{k'g_1} &= \boldsymbol{y}_{kg} \\ \boldsymbol{\mu}_{k''g_2} &= \tilde{\boldsymbol{y}}_{kg} \end{aligned} \quad \text{or} \quad \begin{aligned} \boldsymbol{\mu}_{k'g_1} &= \tilde{\boldsymbol{y}}_{kg} \\ \boldsymbol{\mu}_{k''g_2} &= \boldsymbol{y}_{kg} \end{aligned} \tag{7.13}$$

according to whether $\boldsymbol{y}_{kg}\prec\tilde{\boldsymbol{y}}_{kg}$ or $\tilde{\boldsymbol{y}}_{kg}\prec\boldsymbol{y}_{kg}$; in (7.11-7.13), k' and k'' are values of k depending on the previous values of u_{kg} in the sequential splitting procedure. Regardless of which of (7.11-7.13) is true, the absolute value of the Jacobian is always

$$\det\left[\frac{\partial(\boldsymbol{\mu}_{k'g'},\boldsymbol{\mu}_{k''g''})}{\partial(\boldsymbol{\mu}_{kg},\boldsymbol{y}_{kg})}\right]=\left(1+\frac{p_{kg}^{(1)}}{p_{kg}^{(2)}}\right)^d. \tag{7.14}$$

The relationship of $\boldsymbol{\sigma}_{k'g'}$ and $\boldsymbol{\sigma}_{k''g''}$ to $\boldsymbol{z}_{kg}$ and $\tilde{\boldsymbol{z}}_{kg}$ is the same as (7.11-7.13) above, and hence, it follows that

$$\det\left[\frac{\partial(\boldsymbol{\sigma}_{k'g'},\boldsymbol{\sigma}_{k''g''})}{\partial(\boldsymbol{\sigma}_{kg},\boldsymbol{z}_{kg})}\right]=\left(1+\frac{p_{kg}^{(1)}}{p_{kg}^{(2)}}\right)^d.$$

7.4. *Updating Empty G-components.* For the E-Add move, a empty G-component, g^*, is added to the existing hierarchical mixture model, thus increasing the number of components in the first level by 1 (that is, from G to $G+1$). A uniform $[0,1]$ random variable u_0 is generated and the first level population proportion of g^* is set to u_0. The remaining ω_g are multiplied by $(1-u_0)$ so that the new ω_g's add up to 1. Thus, we have

$$\omega_{g^*}=u_0 \quad \text{and} \quad \omega_g^*=\omega_g(1-u_0)$$

for all $g\neq g^*$. The second level mixture complexity, mixing probabilities, means and variances (namely, K_{g^*}, $\boldsymbol{p}_{g^*}$, $\boldsymbol{\mu}_{g^*}$, and $\boldsymbol{\sigma}_{g^*}$), are generated from the prior specifications (2.1), (2.3) and (2.4). The proposal density corresponding to the E-Add move is

$$q_m(\boldsymbol{x},\boldsymbol{y})=\pi_0(K_{g^*})\,\pi(\boldsymbol{p}_{g^*}\mid K_{g^*})\,\pi(\boldsymbol{\mu}_{g^*}\mid K_{g^*})\,\pi(\boldsymbol{\sigma}_{g^*}\mid K_{g^*})\times(1) \tag{7.15}$$

where π_0 and π are the prior densities given in (2.1-2.4), and (1) is the Jacobian

$$\det\left[\frac{\partial \boldsymbol{y}}{\partial(\boldsymbol{x},\boldsymbol{u})}\right] = (1-u_0)^{(G-1)}. \tag{7.16}$$

The Jacobian takes a relatively simple form since only $\boldsymbol{\omega}$ is affected by re-scaling above; the remainder of the Jacobian is the determinant of an identity matrix. In a similar fashion, the proposal density corresponding to removing an empty G-component is $q_{m'}(\boldsymbol{y},\boldsymbol{x}) = \frac{1}{K_0}$ where K_0 is the number of empty G-components prior to removal. Again, only $\boldsymbol{\omega}$ is affected by this move; ω_gs are updated as $\omega_g^* = \omega_g/(1-\omega_{g^*})$ for all $g \neq g^*$ where g^* is the empty component selected for removal. The Empty-Add and Empty-Remove moves for the K-levels are much simpler than the corresponding G moves. We only have to deal with a single k component as opposed to multiple k components for the G moves. The details of the K moves are hence omitted.

Sarat C. Dass,
Associate Professor,
Department of Statistics & Probability
Michigan State University,
East Lansing, MI 48824
E-mail: sdass@msu.edu
URL: http://www.stt.msu.edu/ sdass

Mingfei Li,
Graduate Student,
Department of Mathematics,
Michigan State University,
East Lansing, MI 48824
E-mail: limingf@msu.edu