The Annals of Applied Statistics
2009, Vol. 3, No. 4, 1448–1466
DOI: 10.1214/09-AOAS266

HIERARCHICAL MIXTURE MODELS FOR ASSESSING FINGERPRINT INDIVIDUALITY[1]

BY SARAT C. DASS AND MINGFEI LI

Michigan State University and Bentley University

The study of fingerprint individuality aims to determine to what extent a fingerprint uniquely identifies an individual. Recent court cases have highlighted the need for measures of fingerprint individuality when a person is identified based on fingerprint evidence. The main challenge in studies of fingerprint individuality is to adequately capture the variability of fingerprint features in a population. In this paper hierarchical mixture models are introduced to infer the extent of individualization. Hierarchical mixtures utilize complementary aspects of mixtures at different levels of the hierarchy. At the first (top) level, a mixture is used to represent homogeneous groups of fingerprints in the population, whereas at the second level, nested mixtures are used as flexible representations of distributions of features from each fingerprint. 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. A Bayesian approach based on reversible jump Markov chain Monte Carlo methodology is developed for the inference of all unknown parameters of hierarchical mixtures. The methodology is illustrated on fingerprint images from the NIST database and is used to make inference on fingerprint individuality estimates from this population.

1. Introduction. Recent court cases have highlighted the need for reporting error rates when an individual is identified based on forensic evidence such as fingerprints. 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 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 have been established, (iii) standards controlling the technique's operation exist and have been maintained, (iv) it has been peer reviewed, and (v) it has a general widespread acceptance [see Pankanti, Prabhakar and Jain (2002) and Zhu, Dass and Jain (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,

Received July 2008; revised June 2009.

[1]Supported by NSF Grant DMS-07-06385.

Key words and phrases. Model-based clustering, Gaussian mixtures, Bayesian inference, reversible jump Markov chain Monte Carlo methods, fingerprint individuality.

fingerprint based identification has been challenged in more than 20 court cases in the United States. To address these concerns, several research investigations have proposed measures that characterize the extent of uniqueness of fingerprints (i.e., fingerprint individuality); see Pankanti, Prabhakar and Jain (2002), Zhu, Dass and Jain (2007) and the references therein. The primary aim of these measures is to capture the inherent variability and uncertainty when an individual is identified based on fingerprint evidence.

The statistical test of hypotheses for fingerprint based identification can be set up as follows: Consider an input fingerprint with an unknown identity I_t being compared to the fingerprint of a claimed identity I_c. The test of hypotheses is

$$H_0 : I_t \neq I_c \quad \text{versus} \quad H_1 : I_t = I_c, \tag{1.1}$$

where H_0 (resp., H_1) is the hypothesis of a negative (resp., positive) identification. The hypotheses posed in the order of negative vs. positive identification (as opposed to the reverse order) allows us to control for the probability of making a false positive identification (i.e., the probability of Type I error). The test of H_0 versus H_1 in (1.1) is carried out by ascertaining the degree of similarity between the two prints and involves two important steps: First, salient fingerprint features are extracted from each print, and second, the collection of features of the two prints are "matched" with each other to obtain the best measure of similarity.

Figure 1 illustrates an example of the feature extraction and matching procedures described in the previous paragraph. Typical fingerprints as in Figure 1 consist of smooth, nonintersecting flow patterns with alternating dark and light lines, called ridges and valleys, respectively. Occasionally, a ridge will either bifurcate or terminate and give rise to an anomaly. The anomalies in the ridge structures are called minutiae which are the fingerprint features used for identifying individuals. Figure 1 shows the locations of minutiae ($x \in R^2$) as white squares for the two fingerprint images extracted using a pattern recognition algorithm described in

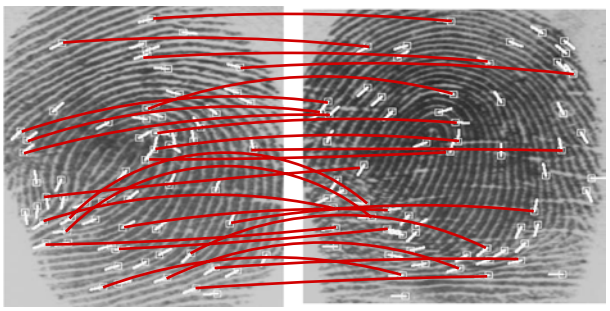


FIG. 1. *Illustrating minutiae matching* [*taken from Pankanti, Prabhakar and Jain* (2002)]. *A total of* $m = 64$ *and* $n = 65$ *minutiae were detected in left and right image, respectively, and* 25 *correspondences* (*i.e., matches*) *were found. The white squares and lines, respectively, represent the minutiae location and the direction of ridge flow at that minutiae.*

Zhu, Dass and Jain (2007). Minutiae information of a fingerprint is easy to extract, permanent (does not change with time) and unique (distinct minutiae patterns for different individuals), making it a popular method for identifying individuals in the forensics community. Subsequently, the number of matches is determined by an optimal rigid transformation that brings the two sets of minutiae as close to each other as possible and counting 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; r_0 is a small prespecified number relative to the size of the fingerprint image. A higher number of matches indicates a higher degree of similarity and favors the rejection of H_0 in (1.1).

The number of matching minutiae in Figure 1 is 25 but the question is: Should H_0 be rejected? In that case, what is the uncertainty or error associated with the decision? This is precisely the issue of fingerprint individuality since error rates associated with the observed match are unknown. Pankanti, Prabhakar and Jain (2002) and Zhu, Dass and Jain (2007) propose using the probability of a random correspondence (PRC) as a measure of fingerprint individuality. Mathematically, the PRC is expressed as

$$\text{PRC}(w|m,n) = P(S \geq w|m,n), \tag{1.2}$$

where the random variable S denotes the number of minutiae matches, w denotes the observed number of matches, and m and n, respectively, are the number of minutiae in the two fingerprint images. The probability in (1.2) is calculated assuming H_0 is true, that is, the pair of prints are impostors coming from two different individuals. Small (resp., large) values of the PRC indicate low (resp., high) levels of uncertainty which correspond to high (resp., low) extent of fingerprint individualization. A PRC of 0.0004, for example, indicates that only 4 out of 10,000 impostor matches will result in matching numbers that are greater than or equal to w. So, having observed w causes us to suspect that H_0 may not be true. The uncertainty associated with this suspicion decreases as the PRC gets smaller (i.e., closer to 0). The connection between the PRC and the hypothesis testing criteria in Daubert (which is one of the five main criteria for the scientific validation of forensic evidence) can be seen as follows: Under the hypotheses testing of (1.1), the PRC is the p-value, computed under H_0, corresponding to the observed number of matches w.

The value of the PRC depends on the distribution of minutiae locations in a pair of prints. Zhu, Dass and Jain (2007) demonstrated that when m and n are large, the distribution of S in (1.2) can be approximated by a Poisson distribution with mean (expected) number of matches

$$\lambda(q_1, q_2, m, n) = mnp(q_1, q_2), \tag{1.3}$$

where $q_h, h = 1, 2$ are the distributions fitted to the minutiae locations in the pair of prints, and $p(q_1, q_2)$ is the probability of a match given by

$$p(q_1, q_2) = \int\int_{(x,y)\,:\,x \in \mathcal{S}(y, r_0)} q_1(x) q_2(y)\, dx\, dy, \tag{1.4}$$

where $x \in R^2$ and $y \in R^2$ are independent minutiae from q_1 and q_2, respectively, and $\mathcal{S}(y, r_0)$ is the square of area $4r_0^2$ centered at y.

The reliability of the PRC computed from a sample of fingerprints depends on (1) how well elicited statistical models fit the distribution of minutiae for different fingerprints, and (2) whether the sample is representative of the target population. The aim in this paper is to develop methodology for (1) while implicitly assuming the validity of (2). Thus, the results in Section 5 are valid for a population which has the fingerprint database as a representative sample.

It is well known, for example, that the distribution of minutiae locations in fingerprints tend to form clusters [see, e.g., Scolve (1979), Stoney and Thornton (1986) and Zhu, Dass and Jain (2007)]. Thus, candidate statistical models have to meet two important requirements: (i) flexibility, that is, the model can represent a variety of minutiae distributions for different fingerprints, and (ii) associated measures of fingerprint individuality can be easily obtained from these models. These considerations led Zhu, Dass and Jain (2007) to propose mixture distributions as candidate choices for q_1 and q_2. Based on mixtures of independent normals, the analytical expression for $p(q_1, q_2)$ in (1.6) becomes

$$p(q_1, q_2) = 4r_0^2 \sum_{k=1}^{K_1} \sum_{k'=1}^{K_2} \prod_{b=1}^{2} \phi_1\big(0 \mid \underbrace{(\mu_{k1}^{(b)} - \mu_{k2}^{(b)})}_{\mu}, \underbrace{(\sigma_{k1}^{(b)})^2 + (\sigma_{k2}^{(b)})^2}_{\sigma^2}\big), \tag{1.5}$$

where $q_h(x) = \sum_{k=1}^{K_h} \prod_{b=1}^{2} \phi_1(x^{(b)} | \mu_{kh}^{(b)}, (\sigma_{kh}^{(b)})^2)$ for $h = 1, 2$, $x = (x^{(1)}, x^{(2)})$ and $\phi_1(\cdot | \mu, \sigma^2)$ is the normal density with mean μ and variance σ^2.

One drawback of Zhu, Dass and Jain (2007) is that no statistical model is elicited on the minutiae for a *population* of fingerprints; standard mixture distributions were proposed for minutiae distributions in each fingerprint separately. As a result, no inference (e.g., confidence intervals) can be obtained for the population version of the PRC. This is our motivation for developing hierarchical mixture models and related inferential tools in this paper. The hierarchical mixture model [see (2.1)] is a model on the minutiae for a population of fingerprints that satisfies both requirements of (i) flexibility and (ii) computational ease mentioned earlier. We assume that the fingerprint population consists of G homogeneous groups with respect to the distribution of minutiae, with q_g and w_g, respectively, denoting the distribution of minutiae locations and population proportion of the gth sub-population, $g = 1, 2, \ldots, G$. For a fingerprint pair coming from the sub-populations g_1 and g_2 with $1 \le g_1, g_2 \le G$, we have $q_1 = q_{g_1}$ and $q_2 = q_{g_2}$ in (1.3). Hence, it follows that the population mean PRC corresponding to w observed matches in the population is given by

$$\overline{\mathrm{PRC}}(w | m, n) = \sum_{g_1=1}^{G} \sum_{g_2=1}^{G} \omega_{g_1} \omega_{g_2} P\big(S \ge w | \lambda(q_{g_1}, q_{g_2}, m, n)\big), \tag{1.6}$$

where S follows a Poisson distribution with mean $\lambda(q_{g_1}, q_{g_2}, m, n)$.

In this paper a Bayesian framework for the inference from hierarchical mixture models is developed, which in turn can be used to make inference for the population mean PRC in (1.6). Hierarchical mixture models contain an unknown number of mixture components at two levels. Green (1995) and Green and Richardson (1997) developed the reversible jump Markov chain Monte Carlo (RJMCMC) approach for estimating the unknown number of mixture components by exploring the space of models of varying dimensions. The RJMCMC procedure developed in this paper generalizes the work of Green and Richardson (1997) to hierarchical mixture models with two levels of hierarchy. The rest of the paper is organized as follows: Section 2 develops hierarchical mixture models for a heterogeneous population of objects (the objects are fingerprints in our application). Sections 3 and 4 develop the Bayesian and RJMCMC framework for inference from hierarchical mixture models. Section 5 discusses the application to fingerprint analysis using PRCs.

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

$$q(\underline{x}) = \sum_{g=1}^{G} \omega_g \prod_{j=1}^{n} q_g(x_j), \tag{2.1}$$

where $\underline{x} = (x_1, x_2, \ldots, x_n)$ are the n observations made on $\mathcal{O}$, ω_g, $g = 1, 2, \ldots, G$ 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 gth cluster given by

$$q_g(x) = \sum_{k=1}^{K_g} p_{kg} f_{kg}(x|\lambda_{kg}), \tag{2.2}$$

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 λ_{kg} denoting the set of all unknown parameters in f_{kg}. Identifiability of the hierarchical mixture model of (2.1) with respect to its components is achieved by imposing the constraints

$$\omega_1 < \omega_2 < \cdots < \omega_G \quad \text{and} \quad \lambda_{1g} \prec \lambda_{2g} \prec \cdots \prec \lambda_{K_g g} \tag{2.3}$$

for each $g = 1, 2, \ldots, G$, where $\prec$ is a partial ordering to be defined later. The set of all unknown parameters in the hierarchical mixture model (2.1) is denoted by $\boldsymbol{\theta} = (G, \boldsymbol{\omega}, \mathbf{K}, \mathbf{p}, \boldsymbol{\lambda})$, where $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_G)$, $\mathbf{K} = (K_1, K_2, \ldots, K_g)$,

$\mathbf{p} = (p_{kg}, k = 1, 2, \ldots, K_g, g = 1, 2, \ldots, G)$ and $\boldsymbol{\lambda} = (\lambda_{kg}, k = 1, 2, \ldots, K_g, g = 1, 2, \ldots, G)$.

Hierarchical mixture models consists of two levels of hierarchy: At the first (top, or G) level, the mixture is used to represent the groups, whereas at the second (or K_g) level, nested mixture models (nested within each $g = 1, 2, \ldots, G$ specification) are used as a flexible representation of the distribution 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. Nonparametric methods were developed in Escobar and West (1995) and Roeder and Wasserman (1997), whereas Ishwaran, James and Sun (2001) and Woo and Sriram (2007) developed methodology for the robust estimation of mixture complexity for count data. As discussed earlier, our approach for estimating mixture complexity will be Bayesian based on the RJMCMC algorithm.

In the subsequent text we assume each f_{kg} is multivariate normal with mean vector $\boldsymbol{\mu}_{kg} \equiv (\mu_{kg}^{(1)}, \mu_{kg}^{(2)}, \ldots, \mu_{kg}^{(d)})' \in R^d$ and covariance matrix $\sum_{kg} \in R^d \times R^d$.

Our analysis on the fingerprint images in the NIST database (see Section 5) reveal that it is adequate to consider diagonal covariance matrices of the form $\sum_{kg} = \operatorname{diag}((\sigma_{kg}^{(1)})^2, (\sigma_{kg}^{(2)})^2, \ldots, (\sigma_{kg}^{(d)})^2)$, where $(\sigma_{kg}^{(b)})^2$ is the variance of the bth component. Four different choices of the covariance matrix $\sum_{kg}$ are considered, namely, diagonal covariance matrix with (i) common entries over k [i.e., $\sigma_{kg}^{(d)} = \sigma_g^{(d)}$, for some common value of $\sigma_g^{(d)}$], (ii) different entries over k, unrestricted covariance matrix with (iii) common entries over k, and (iv) different entries over k. These four choices are evaluated using the Bayes Information Criteria (BIC) which is a model selection criteria that favors parsimonious models consistent with the observed data. The highest BIC was found for the choice of diagonal covariance matrix of (i) or (ii) for almost all of the fingerprints in the NIST database; see Table 1.

Thus, we take the density f_{kg} in (2.2) to be

$$f_{kg}(x|\lambda_{kg}) = \phi_d(x|\boldsymbol{\mu}_{kg}, \boldsymbol{\sigma}_{kg}) = \prod_{b=1}^{d} \phi_1\big(x^{(b)}|\mu_{kg}^{(b)}, (\sigma_{kg}^{(b)})^2\big), \tag{2.4}$$

TABLE 1

Covariance matrix selection: *Entries give the number and percentages of fingerprint images in the NIST database that ranked each covariance model as the top choice based on BIC*

Covariance choice	**(i)**	**(ii)**	**(iii)**	**(iv)**	**Total**
Frequency	1731	238	0	29	1998
Percentage	86.64	11.91	0	1.45	100.00

where $\phi_1(\cdot|\mu, \sigma^2)$ denotes the density of the univariate normal distribution with mean μ and variance σ^2, and $\boldsymbol{\sigma}_{kg} \equiv ((\sigma_{kg}^{(1)})^2, (\sigma_{kg}^{(2)})^2, \ldots, (\sigma_{kg}^{(d)})^2)'$ is the d-variate vector of the variances. The second identifiability condition of (2.3) is re-expressed in terms of the first component of the mean vector as

$$\mu_{1g}^{(1)} < \mu_{2g}^{(1)} < \cdots < \mu_{K_g g}^{(1)}. \tag{2.5}$$

For N independent objects selected randomly from the population, it follows that the distribution of observables for the ith object, $i = 1, 2, \ldots, N$ has the density

$$q(\underline{x}_i) = \sum_{g=1}^{G} \omega_g \prod_{j=1}^{n_i} \sum_{k=1}^{K_g} p_{kg} \phi_d(x_{ij}|\boldsymbol{\mu}_{kg}, \boldsymbol{\sigma}_{kg}), \tag{2.6}$$

where $\underline{x}_i \equiv (x_{ij}, j = 1, 2, \ldots, n_i)$ is the set of n_i observations made on the ith object with each $x_{ij} \in R^d$, for $j = 1, 2, \ldots, n_i$. It follows from independence that the joint distribution of all observables, $\mathbf{x} \equiv (\underline{x}_i, i = 1, 2, \ldots, N)$, from N objects is given by $\prod_{i=1}^{N} q(\underline{x}_i)$.

Two other notations are introduced here: $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ will respectively denote the collection of all $\{\boldsymbol{\mu}_{kg}, k = 1, 2, \ldots, K_g, g = 1, 2, \ldots, G\}$ and $\{\boldsymbol{\sigma}_{kg}, k = 1, 2, \ldots, K_g, g = 1, 2, \ldots, G\}$ vectors. Our goal is to infer the unknown parameters $\boldsymbol{\theta} = (G, \boldsymbol{\omega}, \mathbf{K}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ based on the observed data $\mathbf{x}$.

3. A Bayesian framework for inference. For the subsequent text, some additional notation is introduced. The symbol $I(\mathcal{S})$ will denote the indicator function of the set $\mathcal{S}$, that is, $I(\mathcal{S}) = 1$ if $\mathcal{S}$ is true, and 0, otherwise. The notation $A, B, \ldots|C, D, \ldots$ will denote the distribution of random variables $A, B, \ldots$ conditioned on $C, D, \ldots$, with $\pi(A, B, \ldots|C, D, \ldots)$ denoting the specific form of the conditional distribution. Also, $\pi(A, B, \ldots|\cdot)$ will denote the distribution of $A, B, \ldots$ given the rest of the parameters. We specify a joint prior distribution on $\boldsymbol{\theta}$ in terms of the hierarchical specification

$$\pi(\boldsymbol{\theta}) = \pi(G, \mathbf{K}) \cdot \pi(\boldsymbol{\omega}, \mathbf{p}|G, \mathbf{K}) \cdot \pi(\boldsymbol{\mu}|G, \mathbf{K}) \cdot \pi(\boldsymbol{\sigma}|G, \mathbf{K}). \tag{3.1}$$

The component priors in (3.1) are as follows:

(1) The prior on the mean vector is taken as

$$\begin{aligned} \pi(\boldsymbol{\mu}|\mathbf{K}, G) = \prod_{g=1}^{G} \Bigg[\Bigg(K_g! \prod_{k=1}^{K_g} \phi_1(\mu_{kg}^{(1)}|\mu_0, \tau^2) \Bigg) \\ \times \left(I\left(\mu_{1g}^{(1)} < \mu_{2g}^{(1)} < \cdots < \mu_{K_g g}^{(1)}\right)\right) \\ \times \Bigg(\prod_{b=2}^{d} \prod_{k=1}^{K_g} \phi_1(\mu_{kg}^{(b)}|\mu_0, \tau^2) \Bigg) \Bigg]. \end{aligned} \tag{3.2}$$

The indicator function appears due to the identifiability constraint (2.3) imposed on $\boldsymbol{\mu}$ with resulting normalizing constant $K_g!$ for each $g = 1, 2, \ldots, G$.

(2) The prior distribution of the variances is taken as

$$\pi(\boldsymbol{\sigma}|\mathbf{K}, G) = \prod_{g=1}^{G}\left(\prod_{k=1}^{K_g}\prod_{b=1}^{d} \mathrm{IG}\big((\sigma_{kg}^{(b)})^2|\alpha_0, \beta_0\big)\right), \tag{3.3}$$

where IG denotes the inverse gamma distribution with prior shape and scale parameters α_0 and β_0, respectively.

(3) The prior on the first and second level mixing proportions is taken as

$$\begin{aligned}\pi(\boldsymbol{\omega}, \mathbf{p}|G, \mathbf{K}) &= G! D_G(\boldsymbol{\omega}|\delta_\omega) \cdot I(\omega_1 < \omega_2 < \cdots < \omega_G) \\ &\quad \times \prod_{g=1}^{G} D_{K_g}(\mathbf{p}_g|\delta_p),\end{aligned} \tag{3.4}$$

where $D_H(\cdot|\delta)$ denotes the H-dimensional Dirichlet density with the H-component baseline measure $(\delta, \delta, \ldots, \delta)$, where δ is a prespecified constant, and $\mathbf{p}_g \equiv (p_{1g}, p_{2g}, \ldots, p_{Kg,g})'$. The indicator function arises due to the imposed identifiability constraint (2.3) on $\boldsymbol{\omega}$. It follows that $G!$ is the appropriate normalizing constant for this constrained density, obtained by integrating out $\boldsymbol{\omega}$ and noting that $D_G(\boldsymbol{\omega}|\delta_\omega)$ is invariant under different permutations of $\boldsymbol{\omega}$.

(4) The prior on G and $\mathbf{K}$ is taken as

$$\pi(G, \mathbf{K}) = \pi(G) \cdot \pi(\mathbf{K}|G) = \pi_0(G) \cdot \prod_{g=1}^{G} \pi_0(K_g), \tag{3.5}$$

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

The prior on $\boldsymbol{\theta}$ 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. The reader is referred to our technical report [Dass and Li (2008)] for these specifications.

The likelihood of the hierarchical mixture model 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 hierarchy: (1) The augmented variable $\mathbf{W} \equiv (W_1, W_2, \ldots, W_N)$ denotes the class label of the G sub-populations, that is, $W_i = g$ whenever object i arises from the gth subpopulation, and (2) $\mathbf{Z} \equiv (Z_1, Z_2, \ldots, Z_N)$ with $Z_i \equiv (Z_{ij}, j = 1, 2, \ldots, n_i)$, where $Z_{ij} = k$ for $1 \le k \le K_g$ if x_{ij} arises from the kth mixture component $\phi_d(\cdot|\boldsymbol{\mu}_{kg}, \boldsymbol{\sigma}_{kg})$. We denote the augmented parameter space by the same symbol $\boldsymbol{\theta}$ as before, that is, $\boldsymbol{\theta} =$

$(G, \boldsymbol{\omega}, \mathbf{K}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{W}, \mathbf{Z})$. The augmented likelihood is now

$$\begin{aligned} &\ell(G, \boldsymbol{\omega}, \mathbf{K}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{W}, \mathbf{Z}) \\ &\quad = \prod_{i=1}^{N} \prod_{j=1}^{n_i} \prod_{g=1}^{G} \prod_{k=1}^{K_g} (\phi_d(x_{ij}|\boldsymbol{\mu}_{kg}, \boldsymbol{\sigma}_{kg}))^{I(Z_{ij}=k, W_i=g)}, \end{aligned} \tag{3.6}$$

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

$$\pi(\mathbf{W}, \mathbf{Z}|G, \mathbf{K}, \boldsymbol{\omega}, \mathbf{p}) = \pi(\mathbf{W}|G, \boldsymbol{\omega}) \cdot \pi(\mathbf{Z}|G, \mathbf{K}, \mathbf{W}, \mathbf{p}), \tag{3.7}$$

where $\pi(\mathbf{W}|G, \boldsymbol{\omega}) = \prod_{i=1}^{N} \prod_{g=1}^{G} \omega_g^{I(W_i=g)}$ and

$$\pi(\mathbf{Z}|G, \mathbf{K}, \mathbf{W}, \mathbf{p}) = \prod_{g=1}^{G} \prod_{i\,:\,W_i=g} \prod_{j=1}^{n_i} \prod_{k=1}^{K_g} p_{kg}^{I(Z_{ij}=k)}.$$

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

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{x}) &\propto \ell(G, \boldsymbol{\omega}, \mathbf{K}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{W}, \mathbf{Z}) \times \pi(\mathbf{W}, \mathbf{Z}|G, \mathbf{K}, \boldsymbol{\omega}, \mathbf{p}) \\ &\quad \times \pi(G, \mathbf{K}, \boldsymbol{\omega}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}) \end{aligned} \tag{3.8}$$

based on (3.1), (3.6), (3.7) and observed data $\mathbf{x}$.

4. Posterior inference. The total number of unknown parameters in the hierarchical mixture model depends on the values G and $\mathbf{K}$. Thus, the posterior in (3.8) 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 (3.8) resulting from the hierarchical mixture model specification. We briefly discuss the general RJMCMC implementation here. Let $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ 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 $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^*$ via the proposal distribution $q_m(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$. In order to maintain the time reversibility condition, we require to accept the proposal with probability

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*|\mathbf{x})}{\pi(\boldsymbol{\theta}|\mathbf{x})} \frac{r_{m'} q_{m'}(\boldsymbol{\theta}^*, \boldsymbol{\theta})}{r_m q_m(\boldsymbol{\theta}, \boldsymbol{\theta}^*)}\right\}; \tag{4.1}$$

in (4.1), $q_{m'}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ represents the probability of moving from $\boldsymbol{\theta}^*$ to $\boldsymbol{\theta}$ based on the "reverse" move m', and $\pi(\boldsymbol{\theta}|\mathbf{x})$ denotes the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{x}$. It is crucial that the moves m and m' be reversible [see Green (1995)], meaning

that the densities $q_m(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ and $q_{m'}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ have the same support with respect to a dominating measure. In case $\boldsymbol{\theta}^*$ represents the higher dimensional model, we can first sample $\mathbf{u}$ from a proposal $q_0(\boldsymbol{\theta}, \mathbf{u})$ (with possible dependence on $\boldsymbol{\theta}$), and then obtain $\boldsymbol{\theta}^*$ as a one-to-one function of $(\boldsymbol{\theta}, \mathbf{u})$. In that case, the proposal density $q_m(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ in (4.1) is expressed as

$$q_m(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = q_0(\boldsymbol{\theta}, \mathbf{u}) \Big/ \det\left[\frac{\partial \boldsymbol{\theta}^*}{\partial(\boldsymbol{\theta}, \mathbf{u})}\right], \tag{4.2}$$

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

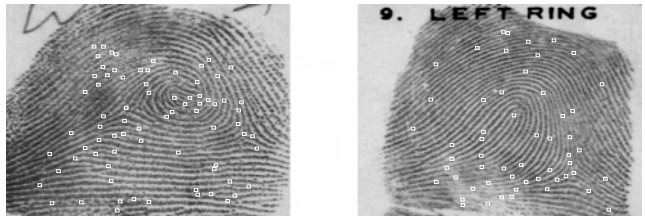
For the inference on hierarchical mixture models, five types of updating steps are considered with reversible pairs of moves, (m, m'), corresponding to moves in spaces of varying dimensions. The outline of the steps are as follows:

$$\begin{cases} (1)\ \text{Update } G \text{ with } (m, m') \equiv (G\text{-split}, G\text{-merge}), \\ (2)\ \text{Update } \mathbf{K}|G, \boldsymbol{\omega}, \mathbf{W} \text{ with } (m, m') \equiv (K\text{-split}, K\text{-merge}), \\ (3)\ \text{Update } \boldsymbol{\omega}|G, \mathbf{K}, \mathbf{W}, \mathbf{Z}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \\ (4)\ \text{Update } \mathbf{W}, \mathbf{Z}|G, \mathbf{K}, \boldsymbol{\omega}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma} \text{ and} \\ (5)\ \text{Update } \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}|G, \mathbf{K}, \boldsymbol{\omega}, \mathbf{W}, \mathbf{Z}. \end{cases} \tag{4.3}$$

Our methodological contribution is the development of the Update G steps (G-split and G-merge) based on a pair of reversible jump moves. The steps for merging and splitting G are described in detail in the Appendix. The Update $\mathbf{K}$ steps are similar to that of Green and Richardson (1997). The other steps (3)–(5) do not involve jumps in spaces of varying dimensions, and can be carried out based on a regular Gibbs proposal. One cycle through steps (1)–(5) completes one iteration of the RJMCMC sampler.

The assessment of convergence of the RJMCMC is carried out based on the methodology of Brooks and Guidici (1998, 2000). A total of 3 chains are run from different starting points and different variance components of the log-likelihood are calculated to obtain 3 diagnostic plots, namely, the plots of (i) the overall and within chain variance, $\hat{V}$ and W_c, (ii) within model and within chain within model variances, W_m and $W_m W_c$, and (iii) between model and between model within chain variances, B_m and $B_m W_c$, against the number of iterations. The merging of the two lines in each plot indicate that the chains have sufficiently mixed.

5. Assessing fingerprint individuality. Our inferential methodology for assessing fingerprint individuality is illustrated using fingerprint images from the NIST Special Database. The NIST fingerprint database is publicly available and consists of 2000 8-bit gray scale fingerprint image pairs of size 512-by-512 pixels. Because of the similarity of the image pairs, only the first image of each pair

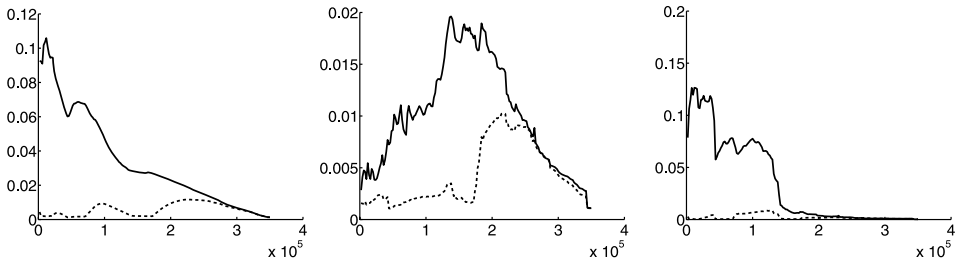


(a) (b)

FIG. 2. *Two fingerprint images from the NIST database with minutiae locations indicated by white squares.*

was used in the statistical modeling. The algorithm described in Zhu, Dass and Jain (2007) (also mentioned in the Introduction) was used to extract minutiae from these images; minutiae could not be automatically extracted from two images of the NIST database due to poor quality and these were discarded from further consideration. Figure 2 shows examples of two fingerprint images from the NIST database with minutiae locations indicated by white squares.

The RJMCMC algorithm developed in the previous section is used to obtain the posterior distribution of $\overline{\text{PRC}}$. The first $N_0 = 100$ fingerprint images from the NIST database are taken as the sample and three chains with starting values obtained using the clustering procedure of Zhu, Dass and Jain (2007) are run. Figure 3 gives the diagnostic plots of the RJMCMC sampler which establish convergence after a burn-in of $B = 250{,}000$ iterations. The posterior distribution of $\overline{\text{PRC}}$ (corresponding to $m = 64$, $n = 65$, $w = 25$ and $r_0 = 15$ pixels) based on 1000 realizations of the RJMCMC after the burn-in period is given in Figure 4 with a posterior mean of 0.6859 and the 95% HPD interval given by $[0.63, 0.735]$. 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 posi-



(a) (b) (c)

FIG. 3. *Convergence diagnostics for the NIST fingerprint database with* $N_0 = 100$. *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.*

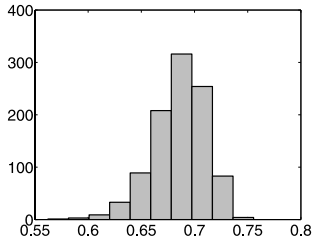


FIG. 4. *Posterior distribution of* $\overline{\text{PRC}}$ *based on* 1000 *realizations of the RJMCMC after* 300,000 *iterations for* $N_0 = 100$.

tive identification. Our analysis actually indicates that the fingerprints in Figure 1 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 result in 25 or more matches.

How many matches does it take to positively identify an individual? Different countries around the world have different standards [Girard (2007)]. In the Netherlands, this number is 12, whereas in South Africa, it is 7. In the United States and the UK, this number is not fixed and depends on expert testimonial. To assess the level of uncertainty associated with these standards, we conduct a study of the PRC based on $w = 7$ matches. The best case scenario corresponds to the combination $(m, n, w) = (7, 7, 7)$ (when all query and template minutiae match with each other) with a mean PRC of 5.09×10^{-5} in Table 2. Note that 7 matches has moderate strength of evidence for declaring a positive match; the PRC implies 5 in 100,000 impostor fingerprint pairs will have all 7 minutiae match with each other. It is also very unlikely that $n = 7$ in real life since fingerprints lifted from a crime scene have far lesser number of minutiae (thus, $m \ll n$) compared to the template it is being matched to. In this latter case, the PRCs are far larger (see Table 2), making the case for positive identification even weaker.

To compare the results of inference using a larger sample size, we ran the RJMCMC sampler for the first $N_0 = 200$ and 500 fingerprint images in the NIST database. The computational complexity increases in two ways: first, it takes longer, on the average, to complete one iteration of the RJMCMC and second, the RJMCMC takes a longer time to converge. On our personal computer with processing speed 2.66 GHz and 1.96 GB of RAM, it took about 12.5, 31.8 and 90.0 hours, respectively, to generate every 50,000 iterations of the RJMCMC

TABLE 2
Mean PRCs for the combinations $(7, 7, n)$

n	7	10	15	55	65	75
Mean PRC	5.09×10^{-5}	1.40×10^{-4}	3.25×10^{-4}	0.0155	0.0333	0.0614

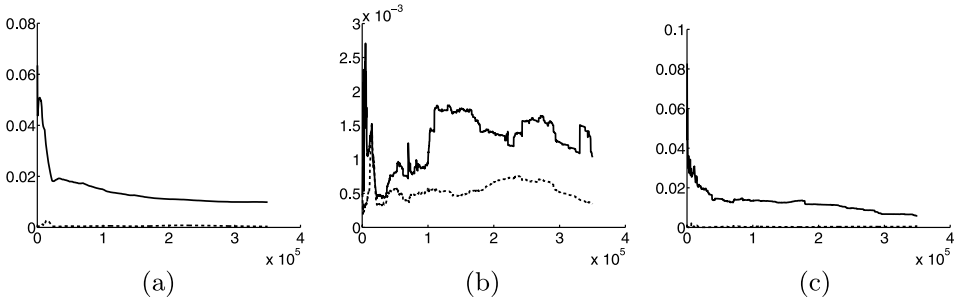


FIG. 5. *Convergence diagnostics for the NIST fingerprint database with* $N_0 = 200$. *Panels* (a), (b) *and* (c), *respectively, show the plots of* $(\hat{V}, W_c)$, $(W_m, W_m W_c)$ *and* $(B_m, B_m W_c)$ *as a function of the iterations. The x-axis unit is* 10,000 *iterations.*

for $N_0 = 100, 200$ and 500. While the RJMCMC converged at 300,000 iterations for $N_0 = 100$, the chain did not converge even at $B = 350{,}000$ iterations for $N_0 = 200$ (see Figure 5) and $N_0 = 500$ (the diagnostic plots are not shown).

Discussion: The RJMCMC sampler is able to accommodate all $N_0 = 1998$ fingerprint images from the NIST database. However, this chain is extremely slow at mixing, and therefore, we do not expect convergence to occur in real time on our computers. Computational demand magnifies exponentially for very large databases such as the US-VISIT program. Thus, for implementation on very large databases, results can be obtained with the help of high-end computing facilities.

One enormous advantage of the methodology outlined in this paper is that the RJMCMC needs to be run on large databases *only once*. After convergence is achieved, inference on $\overline{\text{PRC}}$ for any combination of (w, m, n) can be obtained using formula (1.6). As an illustration based on our smaller sample size of $N_0 = 100$, Table 3 gives the results of this analysis for different combinations of (w, m, n). The entries of Table 3 provides a general guideline to FBI and forensic experts on the extent of uncertainty associated with making a positive identification. Note that when $(w, m, n) = (25, 64, 65)$, the PRC is high, indicating a low extent of individualization. However, Table 3 also provides several combinations of (w, m, n) that favor positive identification with a high degree of individualization. For example, we can look at entries in Table 3 for 95% HPD sets that fall entirely below a threshold, say, T_0. With the choice of $T_0 = 0.003$ (that is, 3 in every 1000 impostor fingerprint pairs will have w or more observed matches), the combinations that allow for positive identification with uncertainty level of at most T_0 are (45, 54, 55), (50, 54, 55), (53, 54, 55), (50, 64, 65) and (53, 64, 65); for these combinations, the probability that the true PRCs occur below T_0 is at least 95%. For larger values of N_0, the size of the HPD sets will decrease due to decreasing variability of the estimate of $\overline{\text{PRC}}$.

In this paper we only considered a two level hierarchical mixture model. 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,

TABLE 3
Posterior means and 95% *HPD sets calculated based on* 1000 *realizations of the RJMCMC for* $N_0 = 100$

w	Mean	HPD
		$(m, n) = (54, 55)$
25	1.33×10^{-1}	$(1.16, 1.53) \times 10^{-1}$
35	1.70×10^{-3}	$(0.80, 3.40) \times 10^{-3}$
45	5.69×10^{-4}	$(0.0007, 2.40) \times 10^{-3}$
50	5.68×10^{-4}	$(0.00002, 2.40) \times 10^{-3}$
53	5.68×10^{-4}	$(0.0002, 7.34) \times 10^{-4}$
		$(m, n) = (64, 65)$
25	6.84×10^{-1}	$(6.27, 7.26) \times 10^{-1}$
35	9.22×10^{-2}	$(0.77, 1.09) \times 10^{-1}$
45	1.90×10^{-3}	$(0.93, 3.60) \times 10^{-3}$
50	6.42×10^{-4}	$(0.0051, 2.40) \times 10^{-3}$
53	5.80×10^{-4}	$(0.0089, 2.40) \times 10^{-3}$
		$(m, n) = (74, 75)$
25	9.50×10^{-1}	$(8.99, 9.87) \times 10^{-1}$
35	6.10×10^{-1}	$(5.54, 6.57) \times 10^{-1}$
45	9.91×10^{-2}	$(0.81, 1.20) \times 10^{-1}$
50	2.12×10^{-2}	$(1.63, 2.73) \times 10^{-2}$
53	7.10×10^{-3}	$(5.10, 9.80) \times 10^{-3}$

individuals form the G groups based on similar characteristics of their 10 fingers, and the distribution of features in each finger is modeled using standard mixtures. Any higher level hierarchical mixture models will be more involved in two ways: (1) The computational costs, including memory and time, since convergence will be much slower to achieve, and (2) the development of reversible moves such as G-merge and G-split for the higher level of mixtures. We are of the view that the best estimate of the population PRC can be obtained if the data is characterized by a model that best represents the way the data is structured and observed. In the case of the US-VISIT, the 3-level hierarchical mixture model is indeed the right way to view the available data. Further research will be needed to see how the computational complexity can be reduced. The availability of high-end computing facilities will definitely be a requirement for fitting higher level hierarchical mixtures.

The central issue for extending the proposed analysis to other biometrics, such as face and iris, is the type of feature extracted for each of the different biometrics. The framework of hierarchical mixture models will apply to these biometric traits but we have to develop mixture models on different feature spaces. The features we used in this paper were minutiae locations, and therefore, we needed mixture models on points in R^2. An additional feature for fingerprints are the minutiae directions (the white lines in Figure 1). In order to run a similar analysis, one would

need to develop suitable mixture models on the product space $R^2 \times [0, 2\pi)$. Similarly, in the case of iris, the feature used is the IrisCode (consisting of a rectangular array of 0s and 1s), and so the statistical models that have to be developed are potentially Markov Random Field models (since there is significant spatial dependence between neighboring 0s and 1s) indexed by a set of parameters. Then, one could postulate that the population consists of G such groups of MRF models. We will also need a distribution for the number of matching features and derive the distribution of this under impostor pairs of IrisCodes.

6. Summary and future work. We have developed Bayesian inference methodology for hierarchical mixture models with application to fingerprint individuality. One way to further reduce the level of uncertainty for a fixed combination (w, m, n) is to increase the number of features used for matching. Our future work will be to derive hierarchical mixture models on the extended feature space consisting of minutiae locations and directions. The challenge here is that the angles are significantly spatially correlated and the minutiae locations exhibit clusters. We are currently developing a model that can account for these minutiae characteristics. We plan to improve our algorithm so that it can be run more quickly on very large databases. 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 modeled by a mixture or a transformation of mixtures.

APPENDIX

In the subsequent text, the identifiability condition (2.5) based on the first components of $\boldsymbol{\mu}_{kg}$ for $k = 1, 2, \ldots, K_g$ will be rewritten using the '$\prec$' symbol as

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

for each $g = 1, 2, \ldots, G$. Let $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ denote two different states of the model space, that is,

$$\begin{aligned} \boldsymbol{\theta} &= (G, \boldsymbol{\omega}, \mathbf{K}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{W}, \mathbf{Z}) \quad \text{and} \\ \boldsymbol{\theta}^* &= (G^*, \boldsymbol{\omega}^*, \mathbf{K}^*, \mathbf{p}^*, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^*, \mathbf{W}^*, \mathbf{Z}^*), \end{aligned} \tag{A.2}$$

where the *s in (A.2) denote a possibly different setting of the parameters.

A.1. 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: Two of the G components, say, g_1 and g_2, with $g_1 < g_2$, are selected randomly for merging into g^* with $\omega_{g^*} = \omega_{g_1} + \omega_{g_2}$.

Step 2: The K-components, K_{g_1} and K_{g_2}, are combined to obtain K_{g^*} in the following way. Adding $K_{g_1} + K_{g_2} = K_t$, we set $K_{g^*} = (K_t + 1)/2$ if K_t is odd, and $K_{g^*} = K_t/2$ if K_t is even.

Step 3: Next, $(\mathbf{p}_{g_1}, \boldsymbol{\mu}_{g_1}, \boldsymbol{\sigma}_{g_1})$ and $(\mathbf{p}_{g_2}, \boldsymbol{\mu}_{g_2}, \boldsymbol{\sigma}_{g_2})$ are merged to obtain $(\mathbf{p}_{g^*}, \boldsymbol{\mu}_{g^*}, \boldsymbol{\sigma}_{g^*})$ as follows. The identifiability conditions of (A.1) hold 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

$$\boldsymbol{\mu}_1 \prec \boldsymbol{\mu}_2 \prec \cdots \prec \boldsymbol{\mu}_{K_t-1} \prec \boldsymbol{\mu}_{K_t} \tag{A.3}$$

with associated probability p_j for $\boldsymbol{\mu}_j$, for $j = 1, 2, \ldots, K_t$. Thus, p_j are a rearrangement of the K_t probabilities in $\mathbf{p}_{g_1}$ and $\mathbf{p}_{g_2}$ according to the partial ordering on $\boldsymbol{\mu}_{g_1}$ and $\boldsymbol{\mu}_{g_2}$ in (A.3). First, the case when K_t is even is considered. Adjacent $\boldsymbol{\mu}$ values in (A.3) are paired

$$\underbrace{\boldsymbol{\mu}_1 \prec \boldsymbol{\mu}_2} \prec \underbrace{\boldsymbol{\mu}_3 \prec \boldsymbol{\mu}_4} \prec \cdots \prec \underbrace{\boldsymbol{\mu}_{K_t-1} \prec \boldsymbol{\mu}_{K_t}} \tag{A.4}$$

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

$$\begin{aligned} \boldsymbol{\mu}^*_{kg^*} &= \frac{p_{2k-1}\boldsymbol{\mu}_{2k-1} + p_{2k}\boldsymbol{\mu}_{2k}}{p_{2k-1} + p_{2k}} \quad \text{and} \\ \boldsymbol{\sigma}^*_{kg^*} &= \frac{p_{2k-1}\boldsymbol{\sigma}_{2k-1} + p_{2k}\boldsymbol{\sigma}_{2k}}{p_{2k-1} + p_{2k}} \end{aligned} \tag{A.5}$$

for $k = 1, 2, \ldots, K_{g^*}$. To obtain $\mathbf{W}^*$ and $\mathbf{Z}^*$, 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. Explicit expressions for the allocation probabilities are provided in Dass and Li (2008). 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, $\{1, 3, 5, \ldots, K_t\}$. 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.

A.2. 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 $\omega_{g_1} = u_0\omega_g$ and $\omega_{g_2} = (1 - u_0)\omega_g$.

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 $\{K_{\min}, K_{\min} + 1, \ldots, K_{\max}\}^2$ 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 $(\mathbf{p}_g, \boldsymbol{\mu}_g, \boldsymbol{\sigma}_g)$ into 2 parts: $(\mathbf{p}_{g_1}, \boldsymbol{\mu}_{g_1}, \boldsymbol{\sigma}_{g_1})$ and $(\mathbf{p}_{g_2}, \boldsymbol{\mu}_{g_2}, \boldsymbol{\sigma}_{g_2})$ such that both $\boldsymbol{\mu}_{g_1}$ and $\boldsymbol{\mu}_{g_2}$ satisfy the constraints (A.1) for $g = g_1$ and g_2. The case of $K_{g_1} + K_{g_2} = 2K_g$ is first considered. A sketch of the split move is best described by the diagram in Figure 6, which introduces the additional variables to be used for performing the split. In Figure 6, $2\mathbf{p}_g$ 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 6 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 $u_{kg}, k = 1, 2, \ldots, K_g$ 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 $h = 0, 2$. The reader is referred to our technical report Dass and Li (2008) for further explanation of these restrictions.

To generate the vector $\underline{u} \equiv (u_{1g}, u_{2g}, \ldots, u_{K_g g})'$, 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 $\underline{v} \equiv (v_{kg}, k = 1, 2, \ldots, K_g)$ is generated to split $2\mathbf{p}_g$ (see Figure 6). Some notation 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

$$p_{kg}^{(1)} = 2v_{kg}p_{kg} \quad \text{and} \quad p_{kg}^{(2)} = 2(1 - v_{kg})p_{kg} \tag{A.6}$$

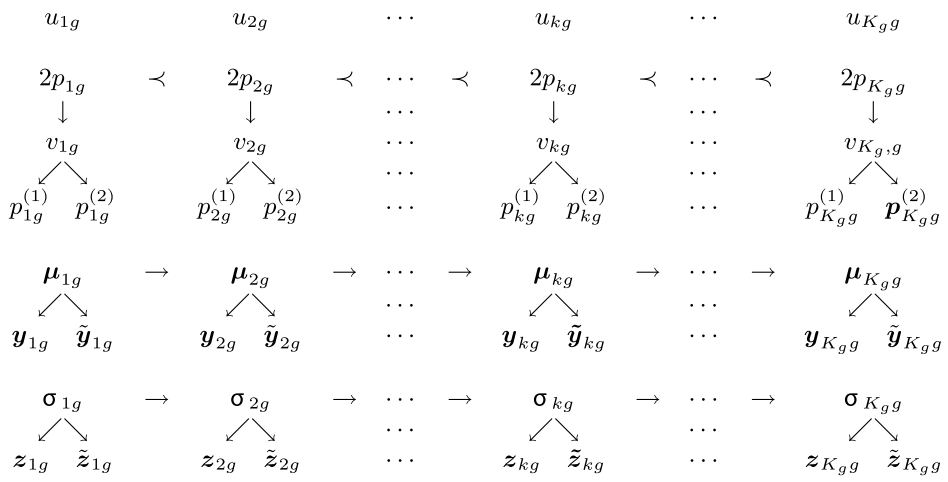


FIG. 6. *Splits of* $2\mathbf{p}_g, \boldsymbol{\mu}_g$ *and* $\boldsymbol{\sigma}_g$. *The partial ordering* $\prec$ *is the ordering on* $\mu_{kg}^{(1)}$*s. The right arrows "*$\rightarrow$*" represents the sequential split for* $\boldsymbol{\mu}_g$ *and* $\boldsymbol{\sigma}_g$.

for $k = 1, 2, \ldots, K_g$, to be the split components from $2p_{kg}$. 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

$$\text{(A.7)} \quad \sum_{k:k\in A_1} p_{kg}^{(1)} + \sum_{k:k\in A_2} 2p_{kg} = 1 \quad \text{and} \quad \sum_{k:k\in A_1} p_{kg}^{(2)} + \sum_{k:k\in A_0} 2p_{kg} = 1$$

for components g_1 and g_2, respectively. The second equation of (A.7) is redundant if the first is assumed since $\sum_{k:k\in A_1} p_{kg}^{(1)} + \sum_{k:k\in A_2} 2p_{kg} + \sum_{k:k\in A_1} p_{kg}^{(2)} + \sum_{k:k\in A_0} 2p_{kg} = 2\sum_{k=1}^{K_g} p_{kg} = 2$. We rewrite the first equation as

$$\text{(A.8)} \qquad \sum_{k:k\in A_1} a_k v_{kg} = 1,$$

where $a_k = 2p_{kg}/(1 - \sum_{k:k\in A_2} 2p_{kg})$. Equation (A.8) implies that the entries of the vector $\underline{v}$ are required to satisfy two restrictions: (1) $0 \le v_{kg} \le 1$ for $k = 1, 2, \ldots, K_g$ from (A.6), and (2) equation (A.8) above. In Dass and Li (2008), an algorithm is given to generate such a $\underline{v}$ where the proposal density can be written down in closed form.

The split of $\boldsymbol{\mu}_g$ and $\boldsymbol{\sigma}_g$ is carried out by generating two new random vectors $\mathbf{y}_{kg}$ and $\mathbf{z}_{kg}$, for $k = 1, 2, \ldots, K_g$; see Figure 6. The generation of $\mathbf{y}_{kg}$ is subject to restrictions arising from constraint (A.1) on $\boldsymbol{\mu}_g$. The other component of the split of $\boldsymbol{\mu}_g$ and $\boldsymbol{\sigma}_g$, $\tilde{\mathbf{y}}_{kg}$ and $\tilde{\mathbf{z}}_{kg}$, are obtained by solving two (vectorized) linear equations [see Dass and Li (2008)]. Our techical report also gives further details of the RJMCMC sampler, including obtaining the new first and second level labels as well as the deriving explicit expressions for the allocation probabilities and the Jacobian of the transformation from $(\boldsymbol{\theta}, \mathbf{u})$ to $\boldsymbol{\theta}^*$.

REFERENCES

BROOKS, S. P. and GIUDICI, P. (1998). Convergence assessment for reversible jump MCMC simulations. In *Bayesian Statistics 6*. Oxford Univ. Press.

BROOKS, S. P. and GIUDICI, P. (2000). Markov chain Monte Carlo convergence assessment via two-way analysis of variance. *J. Comput. Graph. Statist.* **9** 266–285. MR1823805

DASS, S. C. and LI, M. (2008). A Bayesian analysis of hierarchical mixtures with application to clustering fingerprints. Technical Report RM669, Michigan State Univ., Dept. of Statistics & Probability.

DAUBERT V. MERREL DOW PHARMACEUTICALS INC. (1993). 509 U.S. 579, 113 S. Ct. 2786, 125 L.Ed.2d 469.

ESCOBAR, M. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510

GIRARD, J. (2007). *Criminalistics: Forensic Science and Crime*. Jones & Bartlett, Sudbury, MA.

GREEN, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. MR1380810

GREEN, P. and RICHARDSON, S. (1997). On the Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B* **59** 731–792. MR1483213

ISHWARAN, H., JAMES, L. F. and SUN, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *J. Amer. Statist. Assoc.* **96** 1316–1332. MR1946579

NIST SPECIAL DATABASE 4. NIST: 8-bit gray scale images of fingerprint image groups (FIGS). Available at http://www.nist.gov/srd/nistsd4.htm.

PANKANTI, S., PRABHAKAR, S. and JAIN, A. K. (2002). On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24** 1010–1025.

ROEDER, K. and WASSERMAN, L. (1997). Practical Bayesian density estimation using mixtures of normals. *J. Amer. Statist. Assoc.* **92** 894–903. MR1482121

SCOLVE, S. C. (1979). The occurence of fingerprint characteristics as a two dimensional process. *J. Amer. Statist. Assoc.* **74** 588–595. MR0548258

STONEY, D. A. and THORNTON, J. I. (1986). A critical analysis of quantitative fingerprint individuality models. *Journal of Forensic Sciences* **31** 1187–1216.

WOO, M. J. and SRIRAM, T. N. (2007). Robust estimation of mixture complexity for count data. *Comput. Statist. Data Anal.* **51** 4379–4392. MR2364452

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** 391–401.

DEPARTMENT OF STATISTICS & PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824
USA
E-MAIL: sdass@msu.edu
URL: http://www.stt.msu.edu/~sdass

DEPARTMENT OF MATHEMATICAL SCIENCES
BENTLEY UNIVERSITY
175 FOREST STREET
WALTHAM, MASSACHUSETTS 025452
USA
E-MAIL: mli@bentley.edu