SUPPLEMENTARY MATERIAL FOR THE ARTICLE “ASSESSING FINGERPRINT INDIVIDUALITY USING EPIC”

BY CHAE YOUNG LIM AND SARAT C. DASS

Michigan State University

We describe the marked point process models developed for fingerprint minutiae and the details of the Bayesian methodology for inference on the model parameters using a Reversible Jump MCMC method.

4. Spatial Point Processes with Dependent Marks for Fingerprint Minutiae. Let $x_n \equiv \{x_i, i = 1, 2, \cdots, n\}$ denote the collection of $n$ minutiae locations, and for each $x \in x_n$, the minutiae orientation $w_{x}$ denotes the corresponding mark, which takes values in $(0, \pi]$. The distribution of minutiae in a fingerprint image is best described in terms of a hierarchical model involving all random entities. Let $\mathcal{P}(\lambda, h)$ be a marked Poisson process with $\lambda$ and $h$, respectively, denoting the intensity measure and joint density function of marks. The hierarchical model for $(x_n, w_{x_n})$ is given by

\[
\begin{align*}
(\theta, m_{\theta}) &\sim \mathcal{P}(\lambda_1, h_1), \\
(x^{(k)}, w_{x^{(k)}}) | \Phi &\sim \mathcal{P}(\lambda_{2k}, g_k), \text{ for } k = 1, 2, \cdots, K, \text{ and } \\
(x_n, w_{x_n}) &= \bigcup_{k=1}^{K} (x^{(k)}, w_{x^{(k)}}),
\end{align*}
\]

where $K$ is the number of elements in $\theta$; in the above formulation, the intensity measure for $\theta$ is the constant function

\[
\lambda_1(s) = \begin{cases} 
K_0/\text{area}(S), & \text{if } s \in S, \\
0, & \text{otherwise},
\end{cases}
\]

with $S_0$ and $K_0$, respectively, denoting a bounded rectangular region of $R^2$ and a fixed (and known) positive real number. The mark corresponding to $\theta$ is $m_{\theta} \equiv (\gamma, \sigma^2_1, \sigma^2_2, \eta, \rho, \delta^2)$ with density $h_1$ in (2) defined by its component densities

\[
\begin{align*}
\gamma &\sim \text{G}(\alpha_\gamma, \beta_\gamma), \quad \sigma^2_1 \sim \text{IG}(\alpha_1, \beta_1), \quad \sigma^2_2 \sim \text{IG}(\alpha_2, \beta_2), \\
\eta &\sim \text{U}(0, \pi), \quad \rho \sim \text{U}(\rho_{\text{min}}, \rho_{\text{max}}) \quad \text{and} \quad \delta^2 \sim \text{IG}(\alpha_\delta, \beta_\delta),
\end{align*}
\]

independently of each other; in (4) and (5), $\text{G}(\alpha, \beta)$ and $\text{IG}(\alpha, \beta)$ are, respectively, the Gamma and inverse Gamma distributions with shape and scale parameters given by $\alpha$ and $\beta$, and $\text{U}(a, b)$ is the uniform distribution from $a$ to $b$. It is clear from the above specification that the $\gamma$ and
\( \sigma_j^2, j = 1, 2, \) components of \( m_\theta \) should be positive: \( \gamma > 0, \sigma_j^2 > 0 \) for \( j = 1, 2. \) The rest of the components are required to satisfy \( \eta \in (0, \pi], \rho \in (\rho_{\min}, \rho_{\max}), \) and \( \delta^2 > 0. \) The intensity measure \( \lambda_{2k} \) for \( x^{(k)} \) in (2) is

\[
\lambda_{2k}(s) = \begin{cases} \gamma_k \phi_2(s | \theta, \sigma_{1k}^2, \sigma_{2k}^2) & \text{if } s \in S, \\ 0, & \text{otherwise,} \end{cases}
\]

where \( \phi_2(s | \theta, \sigma_{1k}^2, \sigma_{2k}^2) \) is the bivariate normal density function with mean \( \theta \) and diagonal covariance matrix \( \text{diag}(\sigma_{1k}^2, \sigma_{2k}^2). \) The joint density function of marks, \( g_k \) (conditional on \( x^{(k)} \)) is given by

\[
g_k(w | \eta, \rho, \delta_k^2) = V_{n_k} \left( w | \eta, \sum_k (\rho_k, \delta_k^2) \right),
\]

where \( n_k \) is the number of elements in \( x^{(k)} \), \( V_{n_k} \) is the \( n_k \)-variate wrapped normal distribution on \((0, \pi] \) with mean \( \eta_k = (\eta, \eta, \cdots, \eta) \)' \( R^{n_k} \) and covariance matrix \( \sum_k (\rho_k, \delta_k^2) = (\sigma^*_r) \), \( r, s = 1, 2, \cdots, n_k \) with entries given by

\[
\sigma^*_r = \delta^2 \exp(-\rho ||x_r - x_s||),
\]

where \( x_r \) and \( x_s \) are a pair of points from \( x^{(k)} \) and \( || \cdot || \) is the \( R^2 \)-Euclidean norm.

An alternative but equivalent formulation of the hierarchical model (1)-(3) can be described with some additional notation. Denote the class label set \( c_n \equiv \{ c_i, i = 1, 2, \cdots, n \} \) with \( c_i \), corresponding to \( x_i \in x_n \), taking values in the set \( \{ 1, 2, \cdots, K \} \). Also, let \( B_k \equiv \{ x_i : c_i = k \} \), for \( k = 1, 2, \cdots, K \) denote a partition of \( x_n \) which induces a partition of \( (x_n, w_{x_n}) \) into \( K \) sets, given by \( (x_{B_k}, w_{B_k}) \) for \( k = 1, 2, \cdots, K \). The equivalent formulation of (1)-(3) is

\[
(\theta, m_\theta) \equiv \Phi \sim P(\lambda_{1}, h_1),
\]

\[
n | \Phi \sim p_n = \exp(-T) \frac{T^n}{n!}, \]

\[
c_n | n, \Phi \sim \prod_{i=1}^n \gamma_{c_i} D(\theta_{c_i}, \sigma_{1c_i}^2, \sigma_{2c_i}^2) \frac{T}{T}, \]

\[
x_n | c_n, n, \Phi \sim \prod_{i=1}^n \phi_2(x_i | \theta_{c_i}, \sigma_{1c_i}^2, \sigma_{2c_i}^2) D(\theta_{c_i}, \sigma_{1c_i}^2, \sigma_{2c_i}^2), \text{ and}
\]

\[
(w_{x_n} | x_n, c_n, n, \Phi) \sim \prod_{k=1}^K g_k(w_{B_k} | \eta, \rho_k, \delta_k^2),
\]

where \( T \) is defined as

\[
T \equiv \sum_{k=1}^K \gamma_k D(\theta_k, \sigma_{1k}^2, \sigma_{2k}^2),
\]

with

\[
D(\theta_k, \sigma_{1k}^2, \sigma_{2k}^2) = \int_{S_0} \phi_2(s | \theta_k, \sigma_{1k}^2, \sigma_{2k}^2) ds,
\]
and the density $g_k$ is as given in (6). The data augmentation technique for wrapped normal distributions gives rise to the augmented density

$$ (w_x, t_x | x_n, c_n, n, \Phi) \sim \prod_{k=1}^{K} \phi_n \left( w_{B_k} + \pi t_{B_k} \right) \cdot \sum_k (p_k, \delta_k^2) $$

(11)

where $\phi_d (x | \mu, \Sigma)$ is the $d$-variate normal density function with mean $\mu$ and covariance matrix $\Sigma$ and $t_{B_k} = \{ t_x : x \in B_k \}$.

The hierarchical model specification of (7)-(9) and (11) yields the complete (or, augmented) likelihood

$$ \ell(n, c_n, x_n, w_x, t_x | \Phi) = p_n \times \ell_0(c_n | n, \Phi) \times \ell_0(x_n | c_n, n, \Phi) \times \ell_0(w_x, t_x | x_n, c_n, n, \Phi) $$

(12)

where $\ell_0(A, B, \cdots | C, D, \cdots)$ denotes the density of random variables $A, B, \cdots$ conditional on $C, D, \cdots$ given in (7)-(9) and (11). In (12), $\Phi = \{ (\theta_k, \gamma_k, \sigma_{1k}, \sigma_{2k}, \eta_k, \rho_k, \delta_k), k = 1, 2, \cdots, K \}$ denotes the collection of all unknown parameters: $K$ denotes the number of clusters with cluster $k$ having (i) spatial mean $\theta_k$, (ii) spatial variances $\sigma^2_{jk}$ for $j = 1, 2$, (iii) mean of marks $\eta_k$, (iv) covariances between marks governed by the correlation and variance parameters $\rho_k$ and $\delta_k^2$, respectively, and (v) $\gamma_k$ denoting the intensity of the $k$-th cluster which determine the expected total number of points in cluster $k$. For implementing a Bayesian framework of inference, the prior we adopt on $\Phi$ is $\mathcal{P}$ as given in (1). The hyper-parameters in (4) and (5) will be assumed to be fixed and known.

4.1. Posterior Inference. Posterior inference for the likelihood of the hierarchical model in (12) is carried out based on a Markov Chain Monte Carlo (MCMC) algorithm. The MCMC updating steps are (1) update $K$, and for fixed $K$,

1. update $\theta_k, \gamma_k, \sigma^2_{1k}, \sigma^2_{2k}, \eta_k, \rho_k, \delta_k, k = 1, 2, \cdots, K$,
2. update $c_n$, and
3. update $t_x$.

Out of the four updating steps above, only the update of (1) involves parameter spaces of varying dimensions (the remaining updating steps (2)-(4) are regular Metropolis-Hastings (MH) steps). To obtain posterior inference for such spaces of models, Green (1995) and Green and Richardson (1997) developed the Reversible Jump Markov Chain Monte Carlo (RJMCMC) approach for the Bayesian inferential framework. Since their introduction, RJMCMC has been successfully used to obtain posterior inference in many situations. We develop a Reversible Jump Markov Chain Monte Carlo approach to explore the posterior distribution in updating step (1). The updating steps are explained in detail in the Appendix.
The assessment of convergence of the RJMCMC is carried out based on the methodology of Brooks and Guidici (1999, 2000). The diagnostics for assessing convergence utilize the following six quantities: 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 within chain variance, $B_m W_c$. For each monitoring parameter, three figures are obtained: (i) $\hat{V}$ and $W_c$, (ii) $W_m$ and $W_m W_c$, and (iii) $B_m$ and $B_m W_c$ versus the number of iterations. The two plotted lines in each figure should be close to each other to indicate sufficient mixing. Our choice of the monitoring parameter is the logarithm of the complete likelihood (12), namely, $\log \ell(n, c_n, x_n, u_{x_n}, t_{x_n} | \Phi)$, based on the hierarchical model specified by (7)-(9) and (11).

5. Simulation: Convergence plots and tables. Simulation experiments with the spatial domain $S_0 = [0, 100] \times [0, 100]$ and the mark space of $(0, \pi]$ is considered. The results of a simulation experiment with the mark space of $(0, 2\pi]$ is given in the main manuscript. Hyper-parameters in (4) and (5) are set as follows: $\alpha_r$ and $\beta_r$ are derived from $E(\gamma) = 15$ and $Var(\gamma) = 100$. Similarly we set $\alpha_j$ and $\beta_j$ so that $E(\sigma_j^2) = 75$ and $Var(\sigma_j^2) = 10,000$ for $j = 1, 2$. We set $\alpha_3$ and $\beta_3$ so that $E(\delta^2) = 0.1$ and $Var(\delta^2) = 1$. Also, we set $K_0 = 3, (\rho_{\min}, \rho_{\max}) = (0.01, 5), K_{\min} = 2$ and $K_{\max} = 5$. We took the probabilities of selecting move types to be $r_m = r_{m'} = 0.5$ corresponding to the moves $(m, m') = (K-$split, $K-$merge) for $K = K_{\min} + 1, \ldots, K_{\max} - 1$. Also, when $K = K_{\min}, r_m = 1 = 1 - r_{m'}$ and $r_m = 0 = 1 - r_{m'}$ for $K = K_{\max}$. We monitor convergence of $I = 5$ chains with starting values that represent over-dispersion in the chains. The RJMCMC converged after 40,000 iterations (see Figure 1). Table 1 gives the associated statistical inference for the unknown parameters; the true values are compared with the posterior means as well as the 99% credible intervals based on the last 1,000 values from each of the five chains. Figure 2 shows the trace plots of $\hat{V}$ versus $W_c$ for the predictive distribution based on the five chains. It is clear that convergence takes place more rapidly, i.e. by 7,500 iterations.
First, candidates for the

constitutes a model space,

\[ M \]

While the RJMCMC methodology presented here is similar in principle to Richardson and Green (1997), there are some new techniques that we utilize for the associated Bayesian computations. First, candidates for the \( K \)-split move are chosen according to the outcome of a hierarchical

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Mean</th>
<th>Sd</th>
<th>CI</th>
<th>Parameter</th>
<th>True Mean</th>
<th>Sd</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_{11} )</td>
<td>30</td>
<td>33.29</td>
<td>1.82</td>
<td>(32.98, 33.60)</td>
<td>( \theta_{21} )</td>
<td>40</td>
<td>40.99</td>
</tr>
<tr>
<td>( \theta_{12} )</td>
<td>40</td>
<td>40.78</td>
<td>2.45</td>
<td>(38.99, 42.58)</td>
<td>( \theta_{22} )</td>
<td>80</td>
<td>78.10</td>
</tr>
<tr>
<td>( \theta_{13} )</td>
<td>70</td>
<td>71.16</td>
<td>1.84</td>
<td>(69.50, 72.82)</td>
<td>( \theta_{23} )</td>
<td>50</td>
<td>51.15</td>
</tr>
<tr>
<td>( \sigma_{11}^2 )</td>
<td>70</td>
<td>72.34</td>
<td>33.49</td>
<td>(34.38, 246.21)</td>
<td>( \sigma_{21}^2 )</td>
<td>70</td>
<td>61.36</td>
</tr>
<tr>
<td>( \sigma_{12}^2 )</td>
<td>70</td>
<td>98.68</td>
<td>29.09</td>
<td>(46.39, 200.46)</td>
<td>( \sigma_{22}^2 )</td>
<td>70</td>
<td>56.97</td>
</tr>
<tr>
<td>( \sigma_{13}^2 )</td>
<td>70</td>
<td>50.19</td>
<td>23.63</td>
<td>(46.71, 172.01)</td>
<td>( \sigma_{23}^2 )</td>
<td>70</td>
<td>46.27</td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>0.5</td>
<td>0.69</td>
<td>0.20</td>
<td>(0.40, 1.72)</td>
<td>( \delta_1 )</td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>1.5</td>
<td>1.43</td>
<td>0.27</td>
<td>(0.54, 2.10)</td>
<td>( \delta_2 )</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>( \eta_3 )</td>
<td>2.5</td>
<td>2.24</td>
<td>0.11</td>
<td>(1.83, 2.56)</td>
<td>( \delta_3 )</td>
<td>0.17</td>
<td>0.11</td>
</tr>
<tr>
<td>( \rho_1 )</td>
<td>0.15</td>
<td>0.72</td>
<td>0.86</td>
<td>(0.05, 4.74)</td>
<td>( \gamma_1 )</td>
<td>21</td>
<td>22.23</td>
</tr>
<tr>
<td>( \rho_2 )</td>
<td>0.15</td>
<td>0.19</td>
<td>0.30</td>
<td>(0.03, 2.55)</td>
<td>( \gamma_2 )</td>
<td>21</td>
<td>26.71</td>
</tr>
<tr>
<td>( \rho_3 )</td>
<td>0.15</td>
<td>0.93</td>
<td>1.04</td>
<td>(0.07, 4.75)</td>
<td>( \gamma_3 )</td>
<td>28</td>
<td>24.13</td>
</tr>
</tbody>
</table>

**Table 1**
The results of posterior inference based on simulated data for the mark space, \((0, \pi)\).

**Fig 2.** Convergence diagnostics based on predictive characteristics. Panels (a), (b) and (c) show the plot of \((\hat{V}, \hat{W})\), for the proportion of points in region 1, 3 and average marks for region 2, respectively. A solid line represents \(\hat{V}\) and a dotted line represents \(\hat{W}\). The x-axis is taken from 0 to 7,500 to show the initial over-dispersed state. The two lines in each figure coincide all the way up to 50,000 iterations.

**APPENDIX A: THE RJMCMC ALGORITHM**

We refer the reader to Green (1995) and Green and Richardson (1997) for an introduction to the general RJMCMC approach. Based on the likelihood and the prior on \( \Phi \), the posterior distribution for \((\Phi, c_n, t_{x_n})\) is given by (up to a proportionality constant)

\[
\pi_0(\Phi, c_n, t_{x_n} | x_n, w_{x_n}) \propto \ell(n, c_n, x_n, w_{x_n}, t_{x_n} | \Phi) \times \pi_0(\Phi | \Phi_0)
\]

where \( \Phi_0 \) is the collection of hyper-parameters. The collection of all possible values for \((\Phi, c_n, t_{x_n})\) constitutes a model space, \( \mathcal{M} \) say, with varying dimensionality. The posterior distribution is a probability distribution on \( \mathcal{M} \) given the observed data and can be inferred using RJMCMC. While the RJMCMC methodology presented here is similar in principle to Richardson and Green (1997), there are some new techniques that we utilize for the associated Bayesian computations.
clustering algorithm. This approach selects proposals that are highly favored by the observed data, thus increasing the probability of acceptance of such proposals and allowing the chain to mix faster. Second, the split in the angular space cannot be carried out using the linear equations given in Richardson and Green (1997), which are suitable only for variables taking values on the real (or subset of the real) line. A transformation $\tau$ of $(0, \pi]$ which ensures that $\tau(0) = \tau(\pi)$ is proposed for the splitting of angular variables.

We impose an identifiability condition to associate a mixture component with its parameters. We associate component 1 of the mixture with the smallest $\theta_{1k}$ value, component 2 with the second smallest value, and so on, where $\theta_k = (\theta_{1k}, \theta_{2k})$, for $k = 1, 2, \cdots, K$. A re-labelling of the $\theta_k$’s gives

$$\theta_{11} < \theta_{12} < \cdots < \theta_{1K}.$$  \hspace{1cm} (13)

**A.1. Update $K$.** The reversible pair of moves for updating $K$ are $K$-split and $K$-merge. Let $x, y \in \mathcal{M}$ with $x = (\Phi, c_n, t_{n0})$ and $y = (\Phi^*, c_n^*, t_{n0}^*)$, where the *’s denote a possibly different setting of the parameters. When the current state has $k$ components, denote $b_k$ and $d_k$ to be the probabilities of split and merge, respectively, (thus, $b_k = 1 - d_k$ with $d_{K_{\text{min}}} = 0$, $b_{K_{\text{max}}} = 1$ and $b_{K_{\text{max}}} = 0$).

**The $K$-merge move:** The $K$-merge move changes the current $K$ to $K - 1$ (that is, $K^* = K - 1$). Two adjacent components, say $k_1$ and $k_2$ with $k_1 < k_2$ are randomly selected for merging into a new component $k^*$. We merge $(\theta_{k_1}, \gamma_{k_1}, \sigma_{1k_1}^2, \sigma_{2k_1}^2, \eta_{k_1}, \rho_{k_1}, \delta_{k_1}^2)$ and $(\theta_{k_2}, \gamma_{k_2}, \sigma_{1k_2}^2, \sigma_{2k_2}^2, \eta_{k_2}, \rho_{k_2}, \delta_{k_2}^2)$ into $(\theta_{k^*}, \gamma_{k^*}, \sigma_{1k^*}^2, \sigma_{2k^*}^2, \eta_{k^*}, \rho_{k^*}, \delta_{k^*}^2)$ based on the following steps:

- Merge $\gamma_{k_1}$ and $\gamma_{k_2}$ into $\gamma_{k^*}$ by setting $\gamma_{k^*} = \gamma_{k_1} + \gamma_{k_2}$.
- Set $u_1 = \gamma_{k_1} / (\gamma_{k_1} + \gamma_{k_2})$. For the remaining parameters, a generic merging procedure is described. Let $\xi_{k_1}$ and $\xi_{k_2}$ be merged to obtain $\xi_{k^*}$ in the following way:

$$\tau(\xi_{k^*}) = u_1 \tau(\xi_{k_1}) + (1 - u_1)\tau(\xi_{k_2})$$

for some function $\tau$. In the case of variables taking values on the real line, Richardson and Green (1997) take $\tau$ to be the identity function. Thus, the merging of $\theta_{jk_1}$ and $\theta_{jk_2}$ to $\theta_{jk^*}$ for $j = 1, 2, \sigma_{jk_1}^2$ and $\sigma_{jk_2}^2$ to $\sigma_{jk^*}^2$, for $j = 1, 2, \delta_{k_1}^2$ and $\delta_{k_2}^2$ to $\delta_{k^*}^2$, and $\rho_{k_1}$ and $\rho_{k_2}$ to $\rho_{k^*}$ can be carried out with this choice of $\tau$. However, since $\eta$ is an angular variable, we choose $\tau(x) = \min\{x, \pi - x\}$ so that $\tau(0) = \tau(\pi)$. The non-monotonicity of $\tau$ in this case gives rise to two solutions for $\eta_{k^*}$ in (13), say $\eta^{(1)}_{k^*}$ and $\eta^{(2)}_{k^*}$, where without loss of generality, $\eta^{(1)}_{k^*}$ (respectively, $\eta^{(2)}_{k^*}$) is assumed to lie inside (outside of) the line segment joining $\eta_{k_1}$ and $\eta_{k_2}$. The angular distance between $\eta_{k_1}$ and $\eta_{k_2}$ is defined as

$$d(\eta_{k_1}, \eta_{k_2}) = \min\{ |\eta_{k_1} - \eta_{k_2}|, \pi - |\eta_{k_1} - \eta_{k_2}| \}.$$  \hspace{1cm} (14)

We set

$$\eta_{k^*} = \begin{cases} 
\eta^{(1)}_{k^*} & \text{if } d(\eta_{k_1}, \eta_{k_2}) = |\eta_{k_1} - \eta_{k_2}|, \\
\eta^{(2)}_{k^*} & \text{otherwise}.
\end{cases}$$
To obtain $c^*_{n_k}$ and $t^*_{x_n}$, $(x_i, w_i)$ for $i = 1, \cdots, n$ are relabeled so that $B_{k^*} = B_{k_1} \cup B_{k_2}$ and $t_{B_{k^*}} = t_{B_{k_1}} \cup t_{B_{k_2}}$. For the $K$-merge move, the proposal density is given by

$$q_m(x, y) = (K - 1)^{-1}$$

since $K - 1$ is the total number of adjacent pairs that can be merged together.

**The $K$-split move:** The $K$-split move changes the current $K$ to $K + 1$ (thus, $K^* = K + 1$). A candidate component is chosen for splitting, say $k$, with probability $1/K$. Next, $(\theta_k, \gamma_k, \sigma^2_{1k}, \sigma^2_{2k}, \eta_k, \rho_k, \delta^2_k)$ is split into two components, namely,

$$(\theta_k, \gamma_k, \sigma^2_{1k}, \sigma^2_{2k}, \eta_k, \rho_k, \delta^2_k) \quad \text{and} \quad (\theta_{k_1}, \gamma_{k_1}, \sigma^2_{1k_1}, \sigma^2_{2k_1}, \eta_{k_1}, \rho_{k_1}, \delta^2_{k_1}),$$

where $k_1$ and $k_2$ denote the two split components derived from $k$. First, $\gamma_k$ is split into two parts by generating a random variable $u_1$ from $F_1$ in $(0, 1)$ and setting

$$\gamma_{k_1} = u_1 \gamma_k, \quad \text{and} \quad \gamma_{k_2} = (1 - u_1) \gamma_k.$$
where
\[ k_L \equiv k - 1 \text{ and } k_R \equiv k + 1, \text{ and } j1_L \equiv L_j, \text{ and } j1_R \equiv U_j \text{ for } j = 1, 2, \]
with \((L_1, L_2)\) and \((U_1, U_2)\), respectively, denoting the lower left-and upper right-hand corners of the bounded rectangular region \(S_0\). For splitting, generate \(u_2 \sim F_2(a_2, b_2)\) with \(a_2 = \max\{\theta_{1kL}, (\theta_{1k} - (1 - u_1)\theta_{1kR})/u_1\}\) and \(b_2 = \theta_{1kR}\).

- **Split** \(\theta_{2k}\): \(\theta_{2k}\) is split by generating a random variable \(u_3\) from \(F_3(a_3, b_3)\), where \(a_3 = \max\{L_2, (\theta_{2k} - (1 - u_1)u_2)/u_1\}\) and \(b_3 = \min\{u_2, (\theta_{2k} - (1 - u_1)L_2)/u_1\}\).

- **Split** \(\sigma^2_{2k}\) for \(j = 1, 2\): For splitting \(\sigma^2_{1k}\), we require \(u_4 \sim F_4(a_4, b_4)\), where \(a_4 = 0\) and \(b_4 = \sigma^2_{1k}/u_1, \sigma^2_{2k}\) is split similarly via \(u_5 \sim F_5(a_5, b_5)\) where \(a_5 = 0\) and \(b_5 = \sigma^2_{2k}/u_1\).

- **Split** \(\delta^2_k\): The parameter \(\delta^2_k\) is split similarly as \(\sigma^2_k\) via \(u_6 \sim F_6(a_6, b_6)\) with \(a_6 = 0\) and \(b_6 = \delta^2_k/u_1\).

- **Split** \(\rho_k\): \(\rho_k\) is split into two parts via \(u_7 \sim F_7(a_7, b_7)\), where \(a_7 = \max\{\rho_{\min}, (\rho_k - (1 - u_1)\rho_{\max})/u_1\}\) and \(b_7 = \min\{\rho_{\max}, (\rho_k - (1 - u_1)\rho_{\min})/u_1\}\).

- **Split** \(\eta_k\): To split \(\eta_k\), we consider \(\tau(x) = \min\{x, \pi - x\}\). This split is carried out via \(u_8 \sim F_8(a_8, b_8)\), where \(a_8 = \max\{0, (\tau(\eta_k) - (\pi/2)(1 - u_1))/u_1\}\) and \(b_8 = \min\{\pi/2, \tau(\eta_k)/u_1\}\). Two solutions are obtained from (16), and the solution pair \((u_8, \eta_k^{(1)})\) and \((u_8, \eta_k^{(2)})\) that satisfies the distance condition in (14) is selected. If both solution pairs are feasible, then one of them is selected with probability 0.5 each. Let \(u_9\) be the random variable denoting whether \(\eta_k^{(1)}\) or \(\eta_k^{(2)}\) is selected.

To complete the \(K\)-split proposal, we require to obtain the new labels \(c^*_n\) and \(t^*_n\). Recall that
\[
B_k = \{x_i : c_i = k \text{ for } i = 1, \cdots, n\}
\]
is the collection of all points with label \(k\). The new labels for these points are obtained by randomly assigning each \(x_i \in B_k\) to either \(B_{k_1}\) or \(B_{k_2}\) according to the Bayes allocation probabilities
\[
Q_i(k_j) = \frac{\gamma_{k_j}\phi_2(x_i \mid \theta_{k_j}, \sigma^2_{1k_j}, \sigma^2_{2k_j})}{\gamma_{k_1}\phi_2(x_i \mid \theta_{k_1}, \sigma^2_{1k_1}, \sigma^2_{2k_1}) + \gamma_{k_2}\phi_2(x_i \mid \theta_{k_2}, \sigma^2_{1k_2}, \sigma^2_{2k_2})}
\]
for \(j = 1, 2\). It follows that the allocation probability for the \(K\)-split move is
\[
P_{\text{SplitAlloc}} = \prod_{x_i \in B_k} Q_i(k_{ij}),
\]
where \(k_{ij}\) are the realized values of \(k_j\) for \(x_i\), that is, \(c^*_i = k_{ij}\). Once \(c^*_n\) is obtained, \(t^*_x\) is obtained by reallocating \(t_i \in t_{B_k}\) for \(i = 1, \cdots, n_k\) into either \(t_{B_{k_1}}\) or \(t_{B_{k_2}}\) based on the new labels \(c^*_n\).

The proposal density for the \(K\)-split move is given by
\[
q_m(x, y) = (1/K) \times \left( q_0(x, u) \left/ \det \left[ \frac{\partial y}{\partial (x, u)} \right] \right\} \right) \times P_{\text{SplitAlloc}},
\]
where \(q_0(x, u)\) is the density of \(u = (u_1, u_2, \cdots, u_8, u_9)\) given \(x\), and
\[
\det \left[ \frac{\partial y}{\partial (x, u)} \right] = \frac{\gamma_k}{(1 - u_1)^7}
\]
is the absolute value of the Jacobian of the transformation from \((x, u) \to y\).

**Choice of the densities** \(F_j, j = 1, 2, \cdots, 8\): In order to propose good candidates for the \(K\)-split move, we perform hierarchical clustering based on the location and orientation information. The analysis yields two subgroups with labels 1 and 2, say. For the first cluster, we compute the proportion of observations falling in this group, say \(\hat{p}\), location and orientation means, \((\hat{\mu}_x, \hat{\mu}_y)\) and \(\hat{\mu}_m\), and the corresponding variances, \((\hat{\sigma}^2_x, \hat{\sigma}^2_y)\) and \(\hat{\sigma}^2_m\). The density \(F_1\) is taken to be beta\((\alpha_1, \beta_1)\) where \(\alpha_1\) and \(\beta_1\) are chosen so that the mean of \(F_1\) is \(\hat{p}\) and the variance is a small pre-specified value that gives high concentration around \(\hat{p}\). For \(F_2\), we select the normal density with mean \(\hat{\mu}_x\) and a pre-specified standard deviation and restricted inside the interval \((a_2, b_2)\). A similar choice is made for \(F_3\) and \(F_8\). For the variance parameter \(\sigma^2_k\), \(F_4\) is selected as the normal density with mean \(\hat{\sigma}^2_k\) and a pre-specified standard deviation restricted inside the interval \((a_4, b_4)\); \(F_5\) and \(F_6\) is chosen similarly. For \(F_7\), the uniform density on \((a_7, b_7)\) is chosen.

To demonstrate the reversibility of the chain, assume without loss of generality that the move from \(x\) to \(y\) is a move to a higher dimensional space, therefore, representing a \(K\)-split move. The reverse move from \(y\) \(\to x\) represents a \(K\)-merge move to a lower dimensional space. Assume that the value of \(K\) corresponding to state \(x\) is \(k\) (then, it follows that the value of \(K\) for state \(y\) is \(k + 1\)). Reversibility implies and is implied by the fact that the ratio of the transition probabilities

\[
\frac{q_{K\text{split}}(x, y)}{q_{K\text{merge}}(y, x)}
\]

is a well-defined number in \((0, \infty)\). Thus, for reversibility, the ratio cannot be 0, or \(\infty\) or 0/0 form. In other words, if there is a positive transition probability to move from state \(x\) to \(y\), then there should be also a positive probability to move from state \(y\) to \(x\), and vice versa. The \(K\)-split move explicitly describes how to move from a particular state \(x\) to \(y\) with corresponding transition probability

\[
q_{K\text{split}}(x, y) = (1/k) \times \left( q_0(x, u) \right) \left| \frac{\partial y}{\partial(x, u)} \right| \times P_{\text{split Alloc}},
\]

where \(q_0(x, u)\) is the product density of the independent random variables \(u_j \sim F_j(a_j, b_j)\) for \(j = 1, 2, \cdots, 8\), and

\[
\left| \frac{\partial y}{\partial(x, u)} \right| = \frac{\gamma_k}{(1 - u_1)^2}
\]

is the absolute value of the Jacobian of the transformation from \((x, u) \to y\). The \(K\)-merge move is now required to move from \(y\) to \(x\) with positive probability. This is indeed possible if the two components chosen to merge are the very ones that were split in the first place. Executing the steps in the \(K\)-merge move outlined above for \(y\) guarantees that we get back \(x\). Since the value of \(K = (k + 1)\) for \(y\), the transition probability is

\[
q_{K\text{merge}}(y, x) = 1/k.
\]
Thus, the ratio of $q_{\text{Ksplit}}$ to $q_{\text{Kmerge}}$ is a positive number.

A similar argument holds true for the K-merge move $x \rightarrow y$. In this case, the evaluation of $q_{\text{Ksplit}}(y, x)$ which contains expressions of the densities $F_j$, $j = 1, 2, \ldots, 8$ obtained by performing a hierarchical clustering on $y$. Reversibility implies that $q_{\text{Ksplit}}(y, x)$ is positive, which is true since the support of $F_j$ is on the entire interval $(a_j, b_j)$ for $j = 1, 2, \ldots, 8$.

### A.2. Non-dimension changing moves.

**Update $\theta_k$:** For fixed $K$, the parameter $\theta_k \equiv (\theta_{1k}, \theta_{2k})$, for $k = 1, 2, \ldots, K$ will be updated via a MH algorithm subject to the restriction that $L_1 < \theta_{11} < \cdots < \theta_{1K} < U_1$, and $L_2 < \theta_{2k} < U_2$. We update $\theta_k$ in the order determined by the increasing order of variance parameters $\sigma^2_{1k}$, $k = 1, 2, \ldots, K$. Let $m_1, m_2, \ldots, m_K$ be a permutation of indices $1, 2, \ldots, K$ such that $\sigma^2_{1m_1} < \sigma^2_{1m_2} < \cdots < \sigma^2_{1m_K}$. At the $k$-th step, we update $(\theta_{1m_k}, \theta_{2m_k})$ given the remaining parameters; at this point, $(\theta_{1m_1}, \theta_{2m_1})$ for $l = 1, 2, \ldots, (k - 1)$ have been updated while $(\theta_{1m_l}, \theta_{2m_l})$ for $l = (k + 1), (k + 2), \ldots, K$ have not. The following proposal density is used for obtaining a candidate pair, $(\theta^*_{1m_k}, \theta^*_{2m_k})$:

$$
q(\theta^*_{1m_k}, \theta^*_{2m_k} | \cdots) \propto \exp\left(\frac{1}{2} \sum_{j=1}^{2} n_{mk} \left(\frac{\sigma^2 j m_k - \bar{x}_{j m_k}}{\sigma^2_{j m_k}}\right)^2 \times I(\theta^*_{1m_k} \leq \theta_{1m_k} \leq \theta_{1m_k R}) \times I(L_2 \leq \theta^*_{2m_k} \leq U_2),
\right)
$$

where $\bar{x}_{j m_k} = \frac{1}{n_{mk}} \sum_{x_i \in B_{m_k}} x_{ji}$ with $x_i \equiv (x_{1i}, x_{2i})$, and $n_{mk}$ is the number of elements in $B_{m_k}$; in (18), $m_{kL}$ and $m_{kR}$ are the indices from $\{m_1, \ldots, m_{k-1}\}$ such that $\theta^*_{1m_{kL}}$ is the largest value less than $\theta_{1m_k}$ and $\theta^*_{1m_{kR}}$ is the smallest value greater than $\theta_{1m_k}$. If there is no such index, replace the lower and upper bounds by $L_1$ or $U_1$, accordingly. For the proposal density given by (18), the acceptance probability is

$$
\alpha(\theta_{mk}, \theta^*_{mk}) = \min\left\{1, \exp\left(\gamma_{mk}(D(\theta_{mk}, \sigma^2_{1mk}, \sigma^2_{2mk}) - D(\theta^*_{mk}, \sigma^2_{1mk}, \sigma^2_{2mk}))\right)\right\},
$$

where $D$ is as defined in (10).

**Update $\sigma^2_{jk}$ for $j = 1, 2$:** For each $k = 1, 2, \ldots, K$, we consider the following proposal density for a new $\sigma^2_{jk}$ for $j = 1, 2$ in a MH algorithm. The proposal density considered is

$$q(\sigma^2_{jk} | \cdots) \sim IG\left(\alpha_j + n_{jk}/2, \left[1/\beta_j + \frac{1}{2} \sum_{x_i \in B_k} (x_{ji} - \theta_{jk})^2\right]^{-1}\right),$$

where $x_i = (x_{1i}, x_{2i})$ and $\theta_k = (\theta_{1k}, \theta_{2k})$. The acceptance probability is given by

$$
\alpha(\sigma^2_{jk}, \sigma^2_{jk}^*) = \min\left\{1, \exp\left(\gamma_k(D(\theta_k, \sigma^2_{1k}, \sigma^2_{2k}) - D(\theta_k, \sigma^2_{1k}^*, \sigma^2_{2k}^*))\right)\right\},
$$

where $\sigma^2_{j+1k} = \sigma^2_{j+1k}^*$ for $j + 1 \mod 2$.

**Update $\rho_k$:** The conditional posterior distribution of $\rho_k$ given the remaining parameters is given by

$$
\pi(\rho_k | \cdots) \propto \det(\Sigma_{k0})^{-1/2} \exp\left(\frac{1}{2} (\mu_k^\prime \Sigma_{k0}^{-1} \mu_k)\right),
$$

(19)
where \( \mu_k = w_{B_k} + \pi t_{B_k} - \eta_k 1_{n_k}, \Sigma_k = \sum_{B_k}(\rho_k, \delta_k), \Sigma_{k0} = \sum_{B_k}(\rho_k, 1), \) and \( 1_{n_k} \) is the \( n_k \times 1 \) vector with unit entries. To avoid computing the normalizing constant, we discretize a range of \( \rho_k, (\rho_{\min}, \rho_{\max}) \), into subintervals, choose a subinterval with probabilities obtained from evaluating (19) on each subinterval \( \rho_k \), and generate \( \rho_k \) by considering a uniform distribution on the chosen subinterval.

**Updating \( \gamma_k, \eta_k \) and \( \delta_k^2 \):** These updates can be carried out via regular Gibbs sampler (with acceptance probability 1). The conditional posterior distributions of each of the parameters given the remaining parameters are given by:

- \( \pi(\gamma_k | \cdots) \sim G(\alpha_\gamma + n_k, 1/(\beta_\gamma^{-1} + D(\theta_k, \sigma_{1_k}^2, \sigma_{2_k}^2))) \),
- \( \pi(\eta_k | \cdots) \sim \phi_1(\cdot | 1_{n_k}^{-1}(w_{x_{B_k}} + \pi t_{x_{B_k}})/(1_{n_k}^{-1}1_{n_k}), (1_{n_k}^{-1}1_{n_k})^{-1}) \),
- \( \pi(\delta_k^2 | \cdots) \sim IG(\alpha_\delta + n_k/2, [1/\beta_\delta + (1/2)(\mu_k/\Sigma_{k0}^2 \mu_k)]^{-1}) \),

where \( \Sigma_k, \Sigma_{k0} \) and \( \mu_k \) are as defined for updating \( \rho_k \).

**Update \( c_i \):** The update of \( c_i \) given \( c_{-i} \equiv (c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n) \) and other parameters can be carried out via a MH algorithm. The following density is used for proposing a new value of \( c_i^* \):

\[
P(c_i^* = k | c_{-i}, \cdots) \propto \gamma_k \phi_2(x_i | \theta_k, \sigma_{1_k}^2, \sigma_{2_k}^2).
\]

The acceptance probability from using the above proposal density can be obtained as

\[
\alpha(c_i, c_i^*) = \min \left\{ 1, \frac{\prod_{k=1}^K g_{B_k^*}(w_{B_k^*} | \eta_k, \rho_k, \delta_k)}{\prod_{k=1}^K g_{B_k}(w_{B_k} | \eta_k, \rho_k, \delta_k)} \right\},
\]

where \( B_k^* \) and \( B_k \) are as in (17) using \( c_i^* \equiv (c_1, \cdots, c_{i-1}, c_i^*, c_{i+1}, \cdots, c_n) \) and \( c_n \), respectively.

**Update \( t_{x_n} \):** We sequentially update the collection of marks \( t_{B_k} \) from \( k = 1, 2, \cdots, K \). For fixed \( k \), the conditional posterior distribution of each \( t_i \in B_k \) given the rest (namely, \( t_{-i} \equiv (t_1, \cdots, t_{i-1}, t_{i+1}, \cdots, t_n) \)) is given by

\[
\pi(t_i | t_{-i}, \cdots) \propto \phi_{nk}(x_{B_k} | \eta_k, \sum_{B_k}(\rho_k, \delta_k)),
\]

where \( x_{B_k} = w_{B_k} + \pi t_{B_k} \).

**A.3. Updating Empty Components.** The RJMCMC algorithm also incorporates the updating of empty components into the chain, which is done with some modification to the earlier updating \( K \) move types. Empty components can arise naturally when allocating the observations into \( k_1 \) and \( k_2 \) components in the \( K \)-split move type. Instead of rejecting this proposal, we incorporate it into the RJMCMC algorithm by introducing Empty-Add and Empty-Remove move types which are reversible to each other. In the Empty-Add move, an empty \( k^* \) component is added with \((\theta_k^*, \gamma_k^*, \sigma_{1k^*}, \sigma_{2k^*}, \eta_k^*, \rho_k^*, \delta_k^*) \) generated from the prior distribution. The proposal
density corresponding to the Empty-Add move is

\[ q_{m'}(x, y) = \pi(\theta_{k^*}) \pi(\gamma_{k^*} | \alpha_\gamma, \beta_\gamma) \pi(\sigma^2_{1k^*} | \alpha_1, \beta_1) \pi(\sigma^2_{2k^*} | \alpha_2, \beta_2) \]
\[ \times \pi(\eta_{k^*}) \pi(\rho_{k^*}) \pi(\delta^2_{k^*} | \alpha_\delta, \beta_\delta), \]

where \( \pi \)'s are prior densities for the corresponding parameters. In the Empty-Remove move, an empty \( k^* \) component is selected for removal and the corresponding proposal density is \( q_{m'}(x, y) = 1/K_E \), where \( K_E \) is the number of empty \( K \)-components prior to removal.