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 $\boldsymbol{x}_n \equiv \{\, x_i,\, i = 1, 2, \cdots, n \,\}$ denote the collection of n minutiae locations, and for each $x \in \boldsymbol{x}_n$, 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 λ and h, respectively, denoting the intensity measure and joint density function of marks. The hierarchical model for $(\boldsymbol{x}_n, \boldsymbol{w}_{\boldsymbol{x}_n})$ is given by

$$(\boldsymbol{\theta}, \boldsymbol{m}_{\boldsymbol{\theta}}) \equiv \Phi \quad \sim \quad \mathcal{P}(\lambda_1, h_1), \tag{1}$$

$$(\boldsymbol{x}^{(k)}, \boldsymbol{w}^{(k)}_{\boldsymbol{x}^{(k)}}) \,|\, \Phi \quad \overset{ind}{\sim} \quad \mathcal{P}(\lambda_{2k}, g_k), \text{ for } k = 1, 2, \cdots, K, \text{ and} \tag{2}$$

$$(\boldsymbol{x}_n, \boldsymbol{w}_{\boldsymbol{x}_n}) \quad = \quad \bigcup_{k=1}^{K} (\boldsymbol{x}^{(k)}, \boldsymbol{w}^{(k)}_{\boldsymbol{x}^{(k)}}), \tag{3}$$

where K is the number of elements in $\boldsymbol{\theta}$; in the above formulation, the intensity measure for $\boldsymbol{\theta}$ 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 θ is $m_\theta \equiv (\gamma, \sigma_1^2, \sigma_2^2, \eta, \rho, \delta^2)$ with density h_1 in (2) defined by its component densities

$$\gamma \sim \text{G}(\alpha_\gamma, \beta_\gamma), \quad \sigma_1^2 \sim \text{IG}(\alpha_1, \beta_1), \quad \sigma_2^2 \sim \text{IG}(\alpha_2, \beta_2), \tag{4}$$

$$\eta \sim \text{U}(0, \pi), \quad \rho \sim \text{U}(\rho_{\min}, \rho_{\max}) \quad \text{and} \quad \delta^2 \sim \text{IG}(\alpha_\delta, \beta_\delta), \tag{5}$$

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 α and β, and $U(a, b)$ is the uniform distribution from a to b. It is clear from the above specification that the γ and

σ_j^2, $j = 1, 2$, components of m_θ 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 λ_{2k} for $\boldsymbol{x}^{(k)}$ in (2) is

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

where $\phi_2(\, s \,|\, \theta,\, \sigma_1^2,\, \sigma_2^2)$ is the bivariate normal density function with mean θ and diagonal covariance matrix $\text{diag}(\sigma_1^2, \sigma_2^2)$. The joint density function of marks, g_k (conditional on $\boldsymbol{x}^{(k)}$) is given by

$$g_k(\boldsymbol{w} \,|\, \eta_k, \rho_k, \delta_k^2) = V_{n_k}\, \left(\boldsymbol{w} \,|\, \eta_k, \sum\nolimits_k (\rho_k, \delta_k^2)\right), \tag{6}$$

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

$$\sigma_{rs}^* = \delta^2 \, \exp(-\rho ||x_r - x_s||),$$

where x_r and x_s are a pair of points from $\boldsymbol{x}^{(k)}$ 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 $\boldsymbol{c}_n \equiv \{\, c_i,\, i = 1, 2, \cdots, n \,\}$ with c_i, corresponding to $x_i \in \boldsymbol{x}_n$, 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 $\boldsymbol{x}_n$ which induces a partition of $(\boldsymbol{x}_n, \boldsymbol{w}_{\boldsymbol{x}_n})$ into K sets, given by $(\boldsymbol{x}_{B_k}, \boldsymbol{w}_{B_k})$ for $k = 1, 2, \cdots, K$. The equivalent formulation of (1)-(3) is

$$\begin{aligned}
(\boldsymbol{\theta}, \boldsymbol{m_\theta}) \equiv \Phi &\sim \mathcal{P}(\lambda_1, h_1), \\
n \,|\, \Phi &\sim p_n = \exp(-T)\, \frac{T^n}{n!}, && (7) \\
\boldsymbol{c}_n \,|\, n,\, \Phi &\sim \prod_{i=1}^{n} \left[\frac{\gamma_{c_i}\, D(\theta_{c_i}, \sigma_{1c_i}^2, \sigma_{2c_i}^2)}{T}\right], && (8) \\
\boldsymbol{x}_n \,|\, \boldsymbol{c}_n,\, n,\, \Phi &\sim \prod_{i=1}^{n} \left[\frac{\phi_2(\, x_i \,|\, \theta_{c_i},\, \sigma_{1c_i}^2, \sigma_{2c_i}^2)}{D(\theta_{c_i}, \sigma_{1c_i}^2, \sigma_{2c_i}^2)}\right], \text{ and} && (9) \\
(\, \boldsymbol{w}_{\boldsymbol{x}_n} \,|\, \boldsymbol{x}_n,\, \boldsymbol{c}_n,\, n,\, \Phi\,) &\sim \prod_{k=1}^{K} g_k(\boldsymbol{w}_{B_k} \,|\, \eta_k, \rho_k, \delta_k^2),
\end{aligned}$$

where T is defined as

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

with

$$D(\theta_k, \sigma_{1k}^2, \sigma_{2k}^2) \equiv \int_{S_0} \phi_2(\, s \,|\theta_k,\, \sigma_{1k}^2,\, \sigma_{2k}^2)\, ds, \tag{10}$$

and the density g_k is as given in (6). The data augmentation technique for wrapped normal distributions gives rise to the augmented density

$$(\boldsymbol{w}_{\boldsymbol{x}_n}, \boldsymbol{t}_{\boldsymbol{x}_n} \,|\, \boldsymbol{x}_n,\, \boldsymbol{c}_n,\, n,\, \Phi\,) \sim \prod_{k=1}^{K} \phi_{n_k} \left(\boldsymbol{w}_{B_k} + \pi \boldsymbol{t}_{B_k} \,\middle|\, \eta_k,\, \sum\nolimits_k (\rho_k, \delta_k^2) \right) \tag{11}$$

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

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

$$\begin{aligned} \ell(n,\, \boldsymbol{c}_n,\, \boldsymbol{x}_n,\, \boldsymbol{w}_{\boldsymbol{x}_n},\, \boldsymbol{t}_{\boldsymbol{x}_n} \,|\, \Phi\,) \;&=\; p_n \,\times\, \ell_0(\boldsymbol{c}_n \,|\, n,\, \Phi) \,\times\, \ell_0(\boldsymbol{x}_n \,|\, \boldsymbol{c}_n,\, n,\, \Phi) \\ &\times\; \ell_0(\boldsymbol{w}_{\boldsymbol{x}_n}, \boldsymbol{t}_{\boldsymbol{x}_n} \,|\, \boldsymbol{x}_n,\, \boldsymbol{c}_n,\, n,\, \Phi) \end{aligned} \tag{12}$$

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 θ_k, (ii) spatial variances σ_{jk}^2 for $j = 1, 2$, (iii) mean of marks η_k, (iv) covariances between marks governed by the correlation and variance parameters ρ_k and δ_k^2, respectively, and (v) γ_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 Φ 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,

(2) update $(\theta_k, \sigma_{1k}^2, \sigma_{2k}^2, \eta_k, \rho_k, \delta_k), k = 1, 2, \cdots, K$,
(3) update $\boldsymbol{c}_n$, and
(4) update $\boldsymbol{t}_{\boldsymbol{x}_n}$.

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.

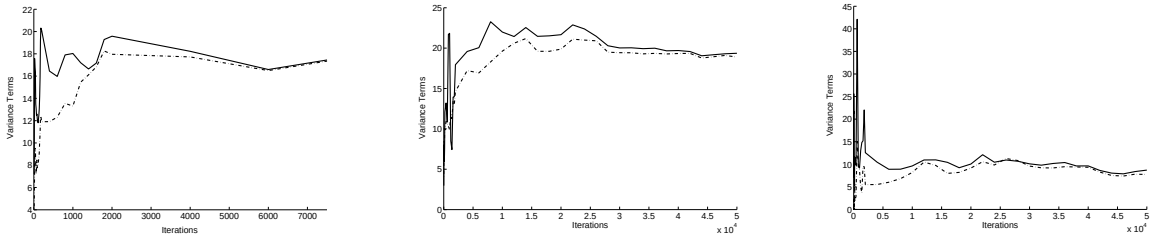


FIG 1. *Convergence diagnostics. Panels (a), (b) and (c) show the plots of* $(\hat{V}, W_c)$, $(W_c, W_m W_c)$ *and* $(B_m, B_m W_c)$, *respectively. A solid line represents the first entry and a dotted line represents the second entry. The x-axis in Panel (a) is taken from* 0 *to* $7,500$ *to show the initial over-dispersion of the chains. The two lines in Panel (a) coincide all the way up to* $50,000$ *iterations.*

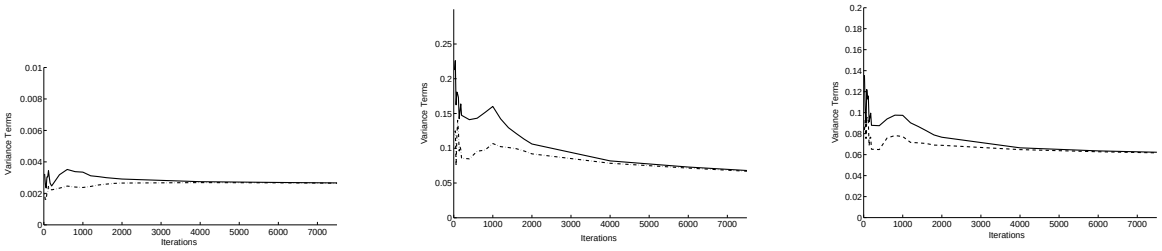
The assessment of convergence of the RJMCMC is carried out based on the methodology of Brooks and Guidici (1999, 2000). The diagnostics for assessing convergence utilize the following six quantities: the overall variance, $\hat{V}$, the within chain variance, W_c, within model variance W_m, within model 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, $\log \ell(\, n, \boldsymbol{c}_n, \boldsymbol{x}_n, \boldsymbol{w}_{\boldsymbol{x}_n}, \boldsymbol{t}_{\boldsymbol{x}_n} \,|\, \Phi\,)$, 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. Hyperparameters in (4) and (5) are set as follows: α_γ and β_γ are derived from $E(\gamma) = 15$ and $Var(\gamma) = 100$. Similarly we set α_j and β_j so that $E(\sigma_j^2) = 75$ and $Var(\sigma_j^2) = 10,000$ for $j = 1, 2$. We set α_δ and β_δ 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 $(m, m') = (K-\text{split}, K-\text{merge})$ for $K = K_{\min} + 1, \cdots, K_{\max} - 1$. 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.

Parameter	True	Mean	Sd	CI	Parameter	True	Mean	Sd	CI
θ_{11}	30	33.29	1.82	$[28.52, 38.27]$	θ_{21}	40	40.99	8.64	$[34.59, 82.01]$
θ_{12}	40	40.78	2.45	$[31.19, 46.05]$	θ_{22}	80	78.10	8.64	$[37.07, 83.75]$
θ_{13}	70	71.16	1.84	$[66.07, 76.52]$	θ_{23}	50	51.15	1.41	$[47.37, 55.23]$
σ^2_{11}	70	72.34	33.49	$[34.38, 246.21]$	σ^2_{21}	70	61.36	17.64	$[31.03, 129.04]$
σ^2_{12}	70	98.68	29.09	$[46.39, 200.46]$	σ^2_{22}	70	56.97	19.30	$[28.23, 141.19]$
σ^2_{13}	70	80.19	23.63	$[46.71, 172.01]$	σ^2_{23}	70	46.27	13.69	$[21.77, 95.47]$
η_1	0.5	0.69	0.20	$[0.40, 1.72]$	δ_1	0.16	0.08	0.05	$[0.04, 0.32]$
η_2	1.5	1.43	0.27	$[0.54, 2.10]$	δ_2	0.16	0.20	0.08	$[0.05, 0.57]$
η_3	2.5	2.24	0.11	$[1.83, 2.56]$	δ_3	0.17	0.11	0.05	$[0.05, 0.32]$
ρ_1	0.15	0.72	0.86	$[0.05, 4.74]$	γ_1	21	22.23	4.48	$[12.32, 35.89]$
ρ_2	0.15	0.19	0.30	$[0.03, 2.55]$	γ_2	21	26.71	4.91	$[15.37, 40.69]$
ρ_3	0.15	0.93	1.04	$[0.07, 4.75]$	γ_3	28	24.13	4.61	$[13.56, 37.44]$

TABLE 1

The results of posterior inference based on simulated data for the mark space, $(0, \pi]$.



(a) (b) (c)

FIG 2. *Convergence diagnostics based on predictive characteristics. Panels (a), (b) and (c) show the plot of* $(\hat{V}, W_c)$*, 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* W_c*. 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 Φ, the posterior distribution for $(\Phi, \boldsymbol{c}_n, \boldsymbol{t}_{\boldsymbol{x}_n})$ is given by (up to a proportionality constant)

$$\pi_0(\Phi, \boldsymbol{c}_n, \boldsymbol{t}_{\boldsymbol{x}_n} \,|\, \boldsymbol{x}_n, \boldsymbol{w}_{\boldsymbol{x}_n}) \propto \ell(n,\, \boldsymbol{c}_n,\, \boldsymbol{x}_n,\, \boldsymbol{w}_{\boldsymbol{x}_n},\, \boldsymbol{t}_{\boldsymbol{x}_n} \,|\, \Phi\,) \times \pi_0(\Phi \,|\, \Phi_0)$$

where Φ_0 is the collection of hyper-parameters. The collection of all possible values for $(\Phi, \boldsymbol{c}_n, \boldsymbol{t}_{\boldsymbol{x}_n})$ 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. First, candidates for the K-split move are chosen according to the outcome of a hierarchical

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 τ 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 θ_{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 θ_k's gives

$$\theta_{11} < \theta_{12} < \cdots < \theta_{1K}. \tag{13}$$

A.1. Update K. The reversible pair of moves for updating K are K-split and K-merge. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{M}$ with $\boldsymbol{x} = (\Phi, \boldsymbol{c}_n, \boldsymbol{t}_{\boldsymbol{x}_n})$ and $\boldsymbol{y} = (\Phi^*, \boldsymbol{c}_n^*, \boldsymbol{t}_{\boldsymbol{x}_n}^*)$, 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 $d_{K_{\min}} = 0$, $b_{K_{\min}} = 1$, $d_{K_{\max}} = 1$ and $b_{K_{\max}} = 0$).

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 $(\theta_{k_1}, \gamma_{k_1}, \sigma^2_{1k_1}, \sigma^2_{2k_1}, \eta_{k_1}, \rho_{k_1}, \delta^2_{k_1})$ and $(\theta_{k_2}, \gamma_{k_2}, \sigma^2_{1k_2}, \sigma^2_{2k_2}, \eta_{k_2}, \rho_{k_2}, \delta^2_{k_2})$ into $(\theta_{k^*}, \gamma_{k^*}, \sigma^2_{1k^*}, \sigma^2_{2k^*}, \eta_{k^*}, \rho_{k^*}, \delta^2_{k^*})$ based on the following steps:

• Merge γ_{k_1} and γ_{k_2} into γ_{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 ξ_{k_1} and ξ_{k_2} be merged to obtain ξ_{k^*} in the following way:

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

for some function τ. In the case of variables taking values on the real line, Richardson and Green (1997) take τ to be the identity function. Thus, the merging of θ_{jk_1} and θ_{jk_2} to θ_{jk^*} for $j = 1, 2$, $\sigma^2_{jk_1}$ and $\sigma^2_{jk_2}$ to $\sigma^2_{jk^*}$ for $j = 1, 2$, $\delta^2_{k_1}$ and $\delta^2_{k_2}$ to $\delta^2_{k^*}$, and ρ_{k_1} and ρ_{k_2} to ρ_{k^*} can be carried out with this choice of τ. However, since η is an angular variable, we choose $\tau(x) = \min\{\, x, \pi - x \,\}$ so that $\tau(0) = \tau(\pi)$. The non-monotonicity of τ in this case gives rise to two solutions for η_{k^*} in (**??**), 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 η_{k_1} and η_{k_2}. The angular distance between η_{k_1} and η_{k_2} is defined as

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

We set

$$\eta_{k^*} = \begin{cases} \eta^{(1)}_{k^*} & \text{if } d(\eta_{k_1}, \eta_{k_2}) = |\eta_{k_1} - \eta_{k_2}|, \text{ and} \\ \eta^{(2)}_{k^*} & \text{otherwise.} \end{cases} \tag{14}$$

To obtain $\boldsymbol{c}_n^*$ and $\boldsymbol{t}_{\boldsymbol{x}_n}^*$, (x_i, w_{x_i}) for $i = 1, \cdots, n$ are relabeled so that $B_{k^*} = B_{k_1} \cup B_{k_2}$ and $\boldsymbol{t}_{B_{k^*}} = \boldsymbol{t}_{B_{k_1}} \cup \boldsymbol{t}_{B_{k_2}}$. For the K-merge move, the proposal density is given by

$$q_{m'}(\boldsymbol{x}, \boldsymbol{y}) = (K-1)^{-1}$$

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, $(\theta_k, \gamma_k, \sigma_{1k}^2, \sigma_{2k}^2, \eta_k, \rho_k, \delta_k^2)$ is split into two components, namely,

$$(\theta_{k_1}, \gamma_{k_1}, \sigma_{1k_1}^2, \sigma_{2k_1}^2, \eta_{k_1}, \rho_{k_1}, \delta_{k_1}^2) \text{ and } (\theta_{k_2}, \gamma_{k_2}, \sigma_{1k_2}^2, \sigma_{2k_2}^2, \eta_{k_2}, \rho_{k_2}, \delta_{k_2}^2),$$

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

$$\gamma_{k_1} = u_1 \gamma_k, \quad \text{and} \quad \gamma_{k_2} = (1-u_1)\gamma_k.$$

We postpone the discussion of the choice of F_1 (as well as the proposal distributions F_j for the random variable u_j for $j = 2, 3, \cdots, 8$ subsequently) until later. The choices are made so that good candidates are generated and the RJMCMC mixes relatively quickly.

A generic split procedure: For splitting the remaining parameters, a split procedure for the generic parameter ξ_k is described. We require to split ξ_k into two components ξ_{k_1} and ξ_{k_2} with the split components satisfying

$$\xi_L \leq \tau(\xi_{k_1}), \tau(\xi_{k_2}) \leq \xi_U, \text{ and } \tau(\xi_k) = u_1 \tau(\xi_{k_1}) + (1-u_1)\tau(\xi_{k_2}) \tag{15}$$

where ξ_L and ξ_U, respectively, are the lower and upper bounds of $\tau(\xi_{k_1})$ and $\tau(\xi_{k_2})$. The function τ is taken as in the K-merge move.

The split can be carried out by generating a random variable $u_* \sim F^*(a_*, b_*)$ where F^* is a density constrained on (a_*, b_*), and setting

$$\xi_{k_1} = u_* \quad \text{and} \quad \xi_{k_2} = \tau^{-1}\left(\frac{\tau(\xi_k) - u_1\, \tau(u_*)}{1-u_1}\right) \tag{16}$$

from the second equality in (15). Note that the first requirement in (15) imposes a lower as well as an upper bound for u^* which we denote by a_* and b_*, respectively. We consider first the splitting of parameters where τ is the identity function.

• **Split** θ_{1k} : Since θ_{1k}, $k = 1, 2, \cdots, K$ satisfy the identifiability constraint (13), we update θ_{1k} sequentially from $k = 1, 2, \cdots, K$. At the k-th step, the split components are given by θ_{1k_1} and θ_{1k_2} satisfying the lower and upper bounds

$$\theta_{1k_L} \leq \theta_{1k_1}, \theta_{1k_2} \leq \theta_{1k_R}$$

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 $u_2 \sim \mathrm{F}_2(a_2, b_2)$ with $a_2 = \max\{\theta_{1k_L}, (\theta_{1k} - (1-u_1)\theta_{1k_R})/u_1\}$ and $b_2 = \theta_{1k_R}$.

• **Split θ_{2k}:** θ_{2k} is split by generating a random variable u_3 from $\mathrm{F}_3(a_3, b_3)$, 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 σ^2_{jk} for $j = 1, 2$:** For splitting σ^2_{1k}, we require $u_4 \sim \mathrm{F}_4(a_4, b_4)$, where $a_4 = 0$ and $b_4 = \sigma^2_{1k}/u_1$. σ^2_{2k} is split similarly via $u_5 \sim \mathrm{F}_5(a_5, b_5)$ where $a_5 = 0$ and $b_5 = \sigma^2_{2k}/u_1$.

• **Split δ^2_k:** The parameter δ^2_k is split similarly as σ^2_{1k} via $u_6 \sim \mathrm{F}_6(a_6, b_6)$ with $a_6 = 0$ and $b_6 = \delta^2_k/u_1$.

• **Split ρ_k:** ρ_k is split into two parts via $u_7 \sim \mathrm{F}_7(a_7, b_7)$, 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 η_k:** To split η_k, we consider $\tau(x) = \min\{x, \pi - x\}$. This split is carried out via $u_8 \sim \mathrm{F}_8(a_8, b_8)$, 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 $\boldsymbol{c}_n^*$ and $\boldsymbol{t}^*_{\boldsymbol{x}_n}$. Recall that

$$B_k = \{x_i \,:\, c_i = k \text{ for } i = 1, \cdots, n\} \tag{17}$$

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

$$\mathcal{Q}_i(k_j) \;=\; \frac{\gamma_{k_j}\phi_2(x_i \,|\, \theta_{k_j}, \sigma^2_{1k_j}, \sigma^2_{2k_j})}{\gamma_{k_1}\phi_2(x_i \,|\, \theta_{k_1}, \sigma^2_{1k_1}, \sigma^2_{2k_1}) + \gamma_{k_2}\phi_2(x_i \,|\, \theta_{k_2}, \sigma^2_{1k_2}, \sigma^2_{2k_2})}$$

for $j = 1, 2$. It follows that the allocation probability for the K-split move is

$$PsplitAlloc = \prod_{x_i \in B_k} \mathcal{Q}_i(k_{ij}),$$

where k_{ij} are the realized values of k_j for x_i, that is, $c_i^* = k_{ij}$. Once $\boldsymbol{c}_n^*$ is obtained, $\boldsymbol{t}^*_{\boldsymbol{x}_n}$ is obtained by reallocating $t_i \in \boldsymbol{t}_{B_k}$ for $i = 1, \cdots, n_k$ into either $\boldsymbol{t}_{B_{k_1}}$ or $\boldsymbol{t}_{B_{k_2}}$ based on the new labels $\boldsymbol{c}_n^*$.

The proposal density for the K-split move is given by

$$q_{m'}(\boldsymbol{x}, \boldsymbol{y}) = (1/K) \times \left(q_0(\boldsymbol{x}, \boldsymbol{u}) \Big/ \det\left[\frac{\partial \boldsymbol{y}}{\partial(\boldsymbol{x}, \boldsymbol{u})}\right] \right) \times PsplitAlloc,$$

where $q_0(\boldsymbol{x}, \boldsymbol{u})$ is the density of $\boldsymbol{u} = (u_1, u_2, \cdots, u_8, u_9)$ given $\boldsymbol{x}$, and

$$\det\left[\frac{\partial \boldsymbol{y}}{\partial(\boldsymbol{x}, \boldsymbol{u})}\right] = \frac{\gamma_k}{(1-u_1)^7}$$

is the absolute value of the Jacobian of the transformation from $(\boldsymbol{x}, \boldsymbol{u}) \to \boldsymbol{y}$.

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, $(\hat{\sigma}_x^2, \hat{\sigma}_y^2)$ and $\hat{\sigma}_m^2$. The density F_1 is taken to be $\texttt{beta}(\alpha_1, \beta_1)$ where α_1 and β_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 σ_{1k}^2, F_4 is selected as the normal density with mean $\hat{\sigma}_x^2$ 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 $\boldsymbol{x}$ to $\boldsymbol{y}$ is a move to a higher dimensional space, therefore, representing a K-split move. The reverse move from $\boldsymbol{y} \to \boldsymbol{x}$ represents a K-merge move to a lower dimensional space. Assume that the value of K corresponding to state $\boldsymbol{x}$ is k (then, it follows that the value of K for state $\boldsymbol{y}$ is $k+1$). Reversibility implies and is implied by the fact that the ratio of the transition probabilities

$$\frac{q_{Ksplit}(\boldsymbol{x}, \boldsymbol{y})}{q_{Kmerge}(\boldsymbol{y}, \boldsymbol{x})}$$

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

$$q_{Ksplit}(\boldsymbol{x}, \boldsymbol{y}) = (1/k) \times \left(q_0(\boldsymbol{x}, \boldsymbol{u}) \Big/ \det\left[\frac{\partial \boldsymbol{y}}{\partial(\boldsymbol{x}, \boldsymbol{u})}\right] \right) \times PsplitAlloc,$$

where $q_0(\boldsymbol{x}, \boldsymbol{u})$ 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

$$\det\left[\frac{\partial \boldsymbol{y}}{\partial(\boldsymbol{x}, \boldsymbol{u})}\right] = \frac{\gamma_k}{(1 - u_1)^7}$$

is the absolute value of the Jacobian of the transformation from $(\boldsymbol{x}, \boldsymbol{u}) \to \boldsymbol{y}$. The K-merge move is now required to move from $\boldsymbol{y}$ to $\boldsymbol{x}$ 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 $\boldsymbol{y}$ *guarantees* that we get back $\boldsymbol{x}$. Since the value of $K = (k + 1)$ for $\boldsymbol{y}$, the transition probability is

$$q_{Kmerge}(\boldsymbol{y}, \boldsymbol{x}) = 1/k.$$

Thus, the ratio of q_{Ksplit} to q_{Kmerge} is a positive number.

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

A.2. Non-dimension changing moves. Update θ_k: For fixed K, the parameter $\theta_k \equiv (\theta_{1k}, \theta_{2k})$, for $k = 1, 2, \cdots, K$ 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 θ_k in the order determined by the increasing order of variance parameters σ^2_{1k}, $k = 1, 2, \cdots, K$. Let $m_1, m_2, \cdots, m_K$ be a permutation of indices $1, 2, \cdots, K$ 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, $(\theta_{1m_l}, \theta_{2m_l})$ for $l = 1, 2, \cdots, (k-1)$ have been updated while $(\theta_{1m_l}, \theta_{2m_l})$ for $l = (k+1), (k+2), \cdots, K$ have not. The following proposal density is used for obtaining a candidate pair, $(\theta^*_{1m_k}, \theta^*_{2m_k})$:

$$\begin{aligned} q(\theta^*_{1m_k}, \theta^*_{2m_k} \mid \cdots) \propto \quad & \exp\left(-\tfrac{1}{2}\textstyle\sum_{j=1}^{2} n_{m_k} \frac{\left(\theta^*_{jm_k} - \bar{x}_{jm_k}\right)^2}{\sigma^2_{jm_k}}\right) \times \\ & I(\theta_{1m_{kL}} \le \theta^*_{1m_k} \le \theta_{1m_{kR}}) \times I(L_2 \le \theta^*_{2m_k} \le U_2), \end{aligned} \tag{18}$$

where $\bar{x}_{jm_k} = \frac{1}{n_{m_k}} \sum_{x_i \in B_{m_k}} x_{ji}$ with $x_i \equiv (x_{1i}, x_{2i})$, and n_{m_k} is the number of elements in B_{m_k}; in (18), m_{kL} and m_{kR} are the indices from $\{m_1, \cdots, m_{k-1}\}$ such that $\theta_{1m_{kL}}$ is the largest value less than θ_{1m_k} and $\theta_{1m_{kR}}$ is the smallest value greater than θ_{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

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

where D is as defined in (10).

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

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

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 $j + 1 \bmod 2$.

Update ρ_k: The conditional posterior distribution of ρ_k given the remaining parameters is given by

$$\pi(\rho_k \mid \cdots) \propto \det(\Sigma_{k0})^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\mu}_k{}' \Sigma_k^{-1} \boldsymbol{\mu}_k)\right), \tag{19}$$

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

Updating γ_k, η_k and δ_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:

- $\pi(\gamma_k \,|\, \cdots) \sim G(\alpha_\gamma + n_k, 1/(\beta_\gamma^{-1} + D(\theta_k, \sigma_{1k}^2, \sigma_{2k}^2)))$,
- $\pi(\eta_k \,|\, \cdots) \sim \phi_1\big(\cdot \;|\, \mathbf{1}_{n_k}'\Sigma_k^{-1}(\boldsymbol{w}_{\boldsymbol{x}_{B_k}} + \pi\, \boldsymbol{t}_{\boldsymbol{x}_{B_k}})/(\mathbf{1}_{n_k}'\Sigma_k^{-1}\mathbf{1}_{n_k})\,,\, (\mathbf{1}_{n_k}'\Sigma_k^{-1}\mathbf{1}_{n_k})^{-1}\big)$,
- $\pi(\delta_k^2 \,|\, \cdots) \sim IG(\alpha_\delta + n_k/2, \big[1/\beta_\delta + (1/2)(\boldsymbol{\mu}_k{}'\Sigma_{k0}^{-1}\boldsymbol{\mu}_k)\big]^{-1})$,
  where Σ_k, Σ_{k0} and $\boldsymbol{\mu}_k$ are as defined for updating ρ_k.

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

$$P(c_i^* = k \,|\, \boldsymbol{c}_{-i}, \cdots) \;\propto\; \gamma_k \phi_2(x_i \,|\, \theta_k,\, \sigma_{1k}^2, \sigma_{2k}^2\,).$$

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

$$\alpha(c_i, c_i^*) = \min\left\{1\,,\, \frac{\prod_{k=1}^K g_{B_k^*}(\boldsymbol{w}_{B_k^*} \,|\, \eta_k, \rho_k, \delta_k)}{\prod_{k=1}^K g_{B_k}(\boldsymbol{w}_{B_k} \,|\, \eta_k, \rho_k, \delta_k)}\right\},$$

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

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

$$\pi(t_i \,|\, \boldsymbol{t}_{-i}, \cdots) \;\propto\; \phi_{n_k}\left(\boldsymbol{x}_{B_k} \;\middle|\; \eta_k,\, \sum\nolimits_{B_k}(\rho_k, \delta_k)\right),$$

where $\boldsymbol{x}_{B_k} = \boldsymbol{w}_{B_k} + \pi\, \boldsymbol{t}_{B_k}$.

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

$$\begin{aligned} q_{m'}(\boldsymbol{x}, \boldsymbol{y}) &= \pi(\theta_{k^*})\,\pi(\gamma_{k^*} \mid \alpha_\gamma, \beta_\gamma)\,\pi(\sigma^2_{1k^*} \mid \alpha_1, \beta_1)\,\pi(\sigma^2_{2k^*} \mid \alpha_2, \beta_2) \\ &\quad \times \pi(\eta_{k^*})\,\pi(\rho_{k^*})\,\pi(\delta^2_{k^*} \mid \alpha_\delta, \beta_\delta), \end{aligned}$$

where πs 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 $q_{m'}(\boldsymbol{x}, \boldsymbol{y}) = 1/K_E$, where K_E is the number of empty K-components prior to removal.

Chae Y. Lim,
Assistant Professor,
Department of Statistics & Probability
Michigan State University,
East Lansing, MI 48824
E-mail: lim@stt.msu.edu
URL: http://www.stt.msu.edu/ sdass

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