Supplementary Material for Repulsive Mixtures

1 Supplementary concerning section 2.4

A possible issue with our repulsive mixture prior is that the full conditionals are nonstandard, complicating posterior computation. To address this, we propose a data augmentation scheme, introducing auxiliary slice variables to facilitate sampling [1]. In particular, letting g_0 be a conjugate prior, introduce a latent variable u which is jointly modeled with γ through

$$\pi(\gamma_1, \ldots, \gamma_k, u) \propto \left(\prod_{h=1}^{k} g_0(\gamma_h) \right) 1\left\{ h(\gamma_1, \ldots, \gamma_k) > u \right\}$$

Here $1(A)$ is the indicator function, equalling 1 if the event A occurs, otherwise being 0. Marginalizing out u, we recover the original density $\pi(\gamma_1, \ldots, \gamma_k)$. This algorithm is a slice sampling algorithm [4], a class of Markov chain Monte Carlo algorithms widely utilized for posterior inference in infinite mixture models [3]. To sample from 1, the derivation of A_j and therefore the invertibility of h is essential. For a repulsion function defined as (4), define $A_j \equiv \bigcap_{\{s:s\neq j\}} [\gamma_j : g\{d(\gamma_s, \gamma_j)\} > u]$. As long as g is invertible in its argument, the set A_j can be calculated, making sampling straightforward. When the repulsion function is defined as (3), one can introduce a latent variable for each product term. Finally, when $m > 1$, the location parameter vector can be sampled element-wise from the truncated distribution. Details on full conditional distributions can be found in the supplementary material. For simplicity, assume that h is defined as (4), g_0 is the m-variate standard normal density and ϕ is the m-variate spherical normal kernel. Let $s_i = 1, \ldots, k$ be the variable indicating which cluster the ith observation belongs to. Let n_j be the number of data points in the jth cluster and let $\bar{y}_j$ be the average of observations in the jth cluster. Let $\alpha_p = (\alpha_1 + n_1, \ldots, \alpha_k + n_k)$. Then the sampling algorithm can be summarized by the following steps:

Step 1. Update s_i for $i = 1, \ldots, n$ by multinomial sampling

$$(s_i|-) \sim Multinomial(l_1, \ldots, l_k), \quad l_j = \frac{p_j \phi(y_i; \gamma_j, \sigma_j I)}{\sum_{h=1}^{k} p_h \phi(y_i; \gamma_h, \sigma_h I)};$$

Step 2. Sample (γ_j, σ_j) for $j = 1, 2, \ldots, k$ from

$$(\gamma_j|-) \sim N\left\{ (1 + n_j/\sigma_j)^{-1} \bar{y}_j n_j/\sigma_j, I(1 + n_j/\sigma_j)^{-1} \right\} 1(\gamma_j \in A_j)$$

$$(1/\sigma_j|-) \sim \text{Ga}\left\{ a_\sigma + \frac{n_j m}{2}, b_\sigma + \frac{1}{2} \sum_{\{i:s_i=j\}} (y_i - \gamma_j)^T (y_i - \gamma_j) \right\}$$

Step 3. Sample u and p from

$$(u|-) \sim Un\left\{0, h(\gamma)\right\}, \quad p \sim Dirichlet\left(\alpha_p\right)$$

2 Supplementary material concerning section 4

Let us provide more details about the densities in figure 2. Density (I) is a Students t density with eight degrees of freedom. Density (IIa) is a two-components mixture of Gaussians with mixture weights $(0{\cdot}3, 0{\cdot}7)$, location parameters $(-0{\cdot}8, 0{\cdot}8)$ and variances $(0{\cdot}2, 0{\cdot}2)$. Density (IIb) is a mixture having the same weights and scale parameters as density (IIa) but location parameters $(-1{\cdot}5, 1{\cdot}5)$, resulting in better separated clusters. Density $(IIIa)$ is a mixture of a Gaussian with mean $0{\cdot}7$, variance $0{\cdot}2$ and weight $0{\cdot}7$ and a Pearson density with mean $-0{\cdot}7$, variance $0{\cdot}2$, weight $0{\cdot}3$, skewness parameter $-0{\cdot}5$ and kurtosis parameter 3. Density $(IIIb)$ is a mixture having the same weights, scale parameters, skewness and kurtosis parameters as density $(IIIa)$ but having location parameter $(-1{\cdot}2, 1{\cdot}2)$, resulting in better separated clusters. Density (IV) is a bivariate mixture of two Gaussians with weight $0{\cdot}5$, location parameters $(0, 0)$ and $(2, 1)$, variances $(0{\cdot}2, 0{\cdot}2)$ and $(0{\cdot}1, 0{\cdot}1)$ and correlation coefficients $0{\cdot}7$ and 0.

Hyperparameters a_σ and b_σ for the density of the scale parameter were set to 3 and 1 respectively. Parameters α_js were all set equal to the same value $\tilde{\alpha}$ and in accordance with [5]'s specification for

the density of the weights. For the non-repulsive model, the kernel locations were given independent standard normal priors. For the repulsive model, we considered a repulsion function defined as (4), with g defined as (5) and we chose g_0 to be the standard normal. Concerning parameters involved in the repulsion function, ν was set equal to 2 while τ was chosen to guarantee a certain separation probability. In particular, the probability used to calibrate τ was chosen to be $0{\cdot}90$ and c in definition 2 was fixed at 2.

The misclassification error was established in terms of divergence between the true similarity matrix and the posterior similarity matrix. Let n be the number of observations, then the similarity matrix is a $(n \times n)$ matrix with (i, j) element equal to one if the jth and the ith observation belong to the same group and zero otherwise. As for the true similarity matrix, observations drawn from the same mixture component were considered as belonging to the same group. Let S be the true similarity matrix and $\hat{S}_h$ be the similarity matrix obtained at the hth Markov chain Monte Carlo iteration. Let $S(i, j)$ be the (i, j) element of S and define m_h as

$$m_h = \sum_{i=1}^{n} \sum_{j=i+1}^{n} 1(\hat{S}_h(i, j) \neq S(i, j))$$

with $1(A)$ being the indicator function, equalling 1 if the event A occurs, otherwise being 0. The misclassification error at the hth iteration was calculated dividing m_h by the total number of pairs of n elements.

An approximation of the Kullback-Leibler divergence at the hth iteration was derived as

$$kl_h = \sum_{j=1}^{s} \log f_0(y_{0j})/f(y_{0j}; \theta_h)$$

with f_0 being the true density, f the fitted density, θ_h the posterior sample at the hth iteration of parameters involved in f and y_{0j} for $i = 1, \ldots, s$ being s draws from the true density f_0. In all simulation examples s was set to $10,000$.

As mentioned in section 4, knowing that the smoothing parameter $\tilde{\alpha}$ directly affects the behavior of the mixture weights, it might be argued that under an accurate choice of $\tilde{\alpha}$, the non-repulsive prior may perform as well as the repulsive prior in emptying the extra components. Hence, we ran the non-repulsive model for different values of $\tilde{\alpha}$. This comparison was done by utilizing dataset (IIb) in figure 2. Table 1 provides posterior summary statistics for parameters involved in the repulsive model and non-repulsive model for different choices of $\tilde{\alpha}$. Clearly, as $\tilde{\alpha}$ decreases, the non-repulsive model empties the extra components. However, we also see that the 95% credible interval of the location parameters now does not include the true value. This might be explained by the fact that as lower values of $\tilde{\alpha}$ are considered, the posterior can concentrate on too few components leading to degenerate results in terms of estimates of specific component parameters.

As mentioned in section 2.4, parameter τ in the repulsion function was calibrated to reach a particular separation probability. In order to assess the sensitivity of results to parameter's calibration, the KL divergence and the sum of extra weights were computed for different choices of separation probability. For this comparison, datasets (IIa) and (IIb) in figure 2 were considered. In practice, at each Markov chain Monte Carlo iteration, an approximation of the KL divergence and the sum of extra weights were obtained. Figure 1 and 2 show the median of these quantities over Markov chain Monte Carlo iterations. Clearly, as the probability of separation among clusters increases, the KL divergence between the fitted density and the true density increases while the sum of extra components decreases. However, as the sample size increase from $n = 100$ to $n = 1000$ the sensitivity of the two quantity of interest to the parameter's choice appears to vanish. Therefore, we expect that as the sample size increases, the choice of the separation probability will not excessively affect results.

3 Assumptions, cited theorems and proofs

3.1 Assumptions

3.1.1 Assumptions B1-B5

Assumptions B1-B5 corresponds to assumptions A1-A5 in [5]. Assumptions differ only in the conditions concerning the prior on the component-specific parameters in assumption A5. In condition

B5, we assume that π is defined as (2) and h is defined as either (3) or (4). For the sake of clarity, let us state assumption B1:

B1) There exists a $q \geq 0$ such that for $\delta_n = (\log n)^q n^{-1/2}$ the following holds

$$\lim_{M\to\infty} \limsup_{n\to\infty} E_n^0 \left\{\Pi\left(\|f - f_0\|_1 \geq M\delta_n | Y_n\right)\right\} = 0$$

3.1.2 Conditions (i), (ii) and (iii) in theorem 3.1 of [6]

(i) The prior on σ has a continuous and positive Lebesgue density ψ on an interval containing σ_0 and its distribution function Ψ, for constants $e_1, e_2, e_3 > 0$, satisfies

$$\Psi(s) \leq \exp(-e_1 s^{-e_2}) \text{ as } s \to 0 \text{ and } 1 - \Psi(s) \leq s^{-e_3} \text{ as } s \to \infty$$

(ii) The prior for the number of components is such that, for constants $d_1, d_2 > 0$,

$$0 < \mu(k) \leq d_1 \exp(-d_2 k) \text{ for all } k \in \mathbb{N}$$

(iii) For each k, the prior for the weights is a Dirichlet with parameters $(\alpha_1, \ldots, \alpha_k)$ such that, for constants a_1, $a_2 > 0$, $a_3 \geq 1$ and for $0 < \epsilon \leq 1/(a_3 k)$ and $j = 1, \ldots, k$

$$a_2 \epsilon^{a_1} \leq \alpha_j \leq a_3$$

3.2 Cited theorems

Theorem 1 ([2]). *Let π_n be a sequence of priors on a class of densities $\mathcal{F}$ equipped with a metric d that can be either the Hellinger or the one induced by the L_1-norm. Assume that for positive sequences $\bar{\epsilon}_n, \tilde{\epsilon}_n \to 0$ such that $n \min(\bar{\epsilon}_n, \tilde{\epsilon}_n) \to \infty$, constants $d_1, d_2, d_3, d_4 > 0$ and sets $\mathcal{F}_n \subseteq \mathcal{F}$, we have*

$$\log D(\bar{\epsilon}_n, \mathcal{F}_n, d) \leq d_1 n \bar{\epsilon}_n^2 \tag{1}$$

$$\pi_n(\mathcal{F} \setminus \mathcal{F}_n) \leq d_3 \exp\left\{-(d_2 + 4) n \tilde{\epsilon}_n^2\right\} \tag{2}$$

$$\pi_n\left\{B_{KL}(f_0; \tilde{\epsilon}_n^2)\right\} \geq d_4 \exp(-d_2 n \tilde{\epsilon}_n^2) \tag{3}$$

where $B_{KL}(f_0; \tilde{\epsilon}_n^2) = \left\{f : \int f_0 \log(f_0/f) \leq \tilde{\epsilon}_n^2; \int f_0 \log(f_0/f)^2 \leq \tilde{\epsilon}_n^2\right\}$.

Then, for $\epsilon_n = \max(\bar{\epsilon}_n, \tilde{\epsilon}_n)$ and a sufficiently large constant $M > 0$, the posterior probability

$$\pi_n\{f : d(f, f_0) > M\epsilon_n | Y_n\} \to 0$$

in P_0^n probability, as $n \to \infty$.

3.3 Proofs

Proof of lemma 1. By assumption B0, $\mu(k = k_0) > 0$. We consider the case f is a finite mixture with k_0 components. By assumption A1, for each $\eta > 0$ there is a corresponding $\delta > 0$ such that, for any given $y \in \mathcal{Y}$ and for all $\gamma_1, \gamma_2 \in \Gamma$ with $|\gamma_1 - \gamma_2| < \delta$, we have that $|\phi(y; \gamma_1) - \phi(y; \gamma_2)| < \eta$. Let $S_\delta = \Gamma_\delta \times P_\delta$ with $\Gamma_\delta = \{\gamma : |\gamma_j - \gamma_{0j}| \leq \delta, j \leq k_0\}$ and $P_\delta = \{p : |p_j - p_{0j}| \leq \delta, j \leq k_0\}$. By assumption A1 and A2, for any given y and for any $\eta > 0$, there is a $\delta > 0$ such that $|f_0 - f| \leq \eta$ if $\theta \in S_\delta$. This means that, $f \to f_0$ as $\theta \to \theta_0$, for any given y. Equivalently, we can say that $|\log(f_0/f)| \to 0$ pointwise as $\theta \to \theta_0$. Notice that

$$|\log(f_0/f)| \leq \left|\log\left\{\sup_{\gamma\in D_0} \phi(\gamma)\right\} - \log\left\{\inf_{\gamma\in D_0} \phi(\gamma)\right\}\right|$$

By assumption A3 and applying the dominated convergence theorem, for any $\epsilon > 0$ there is a $\delta > 0$ such that $\int f_0 \log(f_0/f) < \epsilon$ if $\theta \in S_\delta$. By the independence of the weights and the parameters of the kernel,

$$\Pi(KL(f_0, f) < \epsilon) \geq \lambda(P_\delta)\pi(\Gamma_\delta)$$

Assumption A4 combined with the fact that $\{\gamma : ||\gamma - \gamma_0||_1 \leq \delta\} \subseteq \Gamma_\delta$ result in $\pi(\Gamma_\delta) > 0$. Finally, since $\lambda = Dirichlet(\alpha)$, it can be shown that $\lambda(P_\delta) > 0$. □

Proof of lemma 2. Let $D = \{\gamma : ||\gamma - x||_1 < \upsilon/2\}$. By the assumptions on h, given a vector x satisfying condition A4 in lemma 2, $h(\gamma) > 0$ for γ such that $d(\gamma_s, x_s) < \upsilon/2$ for $s = 1, \ldots, k$. Since,

$$D \subseteq \{\gamma : d(\gamma_s, x_s) < \upsilon/2; s = 1, \ldots, k\},$$

it follows that $h(\gamma) > 0$ on D. By assumption, g_0 is positive on Γ, therefore it follows that $\pi(\gamma) > 0$ on D. □

Proof of lemma 3. To prove lemma 3 we need to show that the three conditions of theorem 2.1 in [2] are satisfied. First, define $D(\epsilon, \mathcal{F}, d_s)$ as the maximum number of points in $\mathcal{F}$ such that the distance, with respect to metric d_s, between each pair is at least ϵ. Let d_s be either the Hellinger metric or the one induced by the L1-norm. For given sequences $k_n, a_n, u_n \uparrow \infty$ and $b_n \downarrow 0$ define

$$\mathcal{F}_n^{(k)} = \left\{ f : f = \sum_{j=1}^{k} p_j \phi(\gamma_j, \sigma), \gamma \in (-a_n, a_n)^k, \sigma \in (b_n, u_n) \right\}$$

and $\mathcal{F}_n = \cup_{j=1}^{k_n} \mathcal{F}_n^{(j)}$. As it is shown in [6], for constants $f_2 \geq f_1 > 0$ and $l_1, l_2, l_3 > 0$, derived below to satisfy condition (2) and (3) in [2], and defining $f_1 \log n \leq k_n \leq f_2 \log n$, $a_n = l_3 \left(\log \bar{\epsilon}_n^{-1}\right)^{1/2}$, $b_n = l_1 (\log \bar{\epsilon}_n^{-1})^{-1/e_2}$ and $u_n = \bar{\epsilon}_n^{-l_2}$, $\log D(\bar{\epsilon}_n, \mathcal{F}_n, d_s) \lesssim n\bar{\epsilon}_n^2$ with $\bar{\epsilon}_n = n^{-1/2} \log n$.

Let $A_{n,j} = (-a_n, a_n)^j$. In order to show condition (2) of theorem 2.1. in [2], we need to show that there is a constant $q_1 > 0$ such that $\pi(A_{n,k}^C) \lesssim \exp(-q_1 a_n^2)$. From the exchangeability assumption it follows

$$\begin{aligned} pr(A_{n,k}^C | k = s) &= \textstyle\sum_{j=1}^{s} \frac{s!}{j!(s-j)!} \pi\left(A_{n,j}^C \times A_{n,s-j}\right) \\ &\leq s \textstyle\sum_{j=1}^{s} \frac{(s-1)!}{(j-1)!(s-j)!} \pi\left(A_{n,j}^C \times A_{n,s-j}\right) \leq s \pi_m(A_{n,1}^C) \end{aligned}$$

Therefore, condition C1 implies that, for a positive constant q_1 we have $\pi(A_{n,k}^C) \lesssim E(k) \exp(-q_1 a_n^2)$ with $E(k) < \infty$ by condition (ii). Given a positive constant z_2 chosen to satisfy condition (3) in theorem 2.1 of [2], let $f_1 \geq (z_2+4)/d_2$, $l_1 \leq \{e_1/4(z_2+4)\}^{1/e_2}$, $l_2 \geq 4(z_2+4)/e_3$ and $l_3 \geq \{4(z_2+4)/q_1\}^{1/2}$. Under these values of f_1, l_1, l_2 and l_3, following [6], assumptions (i), (ii) and assumption C1 imply $\Pi(\mathcal{F} \setminus \mathcal{F}_n) \lesssim \exp\left\{-(z_2+4) n\tilde{\epsilon}_n^2\right\}$ with $\tilde{\epsilon}_n = n^{-1/2}(\log n)^{1/2}$.

To show condition (3) of theorem 2.1 in [2], we can again follow the proof of theorem 3.1. in [6]. The only thing we need to show is that, there are constants $u_1, u_2, u_3 > 0$ such that for any $\epsilon_n \leq u_3$

$$\pi(||\gamma - \gamma_0||_1 \leq \epsilon_n) \geq u_1 \exp\left\{-u_2 k_0 \log(1/\epsilon_n)\right\}$$

that is guaranteed by condition C2. Therefore, it can be easily showed that, for sufficiently large n, $z_2 > 0$ and $\tilde{\epsilon}_n = n^{-1/2}(\log n)^{1/2}$, $\Pi\left\{B_{KL}(f_0, \tilde{\epsilon}_n^2)\right\} \gtrsim \exp(-z_2 n\tilde{\epsilon}_n^2)$. □

Proof of lemma 4. First, let us check that condition C1 is satisfied. Clearly, under the assumptions on h, π leads to exchangeable atoms. Under the assumptions on π, the following holds

$$\pi_m(|\gamma_1| \geq t) = \int_{|\gamma_1| \geq t} \pi_m(\gamma_1) d\gamma_1 \leq c_1 c_2 \int_{|\gamma_1| \geq t} g_0(\gamma_1) d\gamma_1$$

with c_1 and c_2 defined as in (2). It follows that there exists a constant $n_1 > 0$ such that $\pi_m(|\gamma_1| \geq t) \lesssim \exp(-n_1 t^2)$.

Now let us verify condition C2. Assumptions on h imply that for any $0 < \epsilon < 1$ there is a corresponding $0 < \delta = g^{-1}(\epsilon)$ and constants $w_1 > 0$ such that $h(\gamma) \geq w_1 \epsilon^{k_0}$ for all γ satisfying $\min_{\{(s,j):s<j\}} d(\gamma_j, \gamma_s) \geq \delta$. Let u_3 be defined as

$$u_3 = \min\left[\epsilon_1/2, g\left(\delta_1\right)\right]$$

with ϵ_1 defined as in assumption B0 and $\delta_1 = \epsilon_1(1 - 1/k_0)$. By assumption $\epsilon < u_3$ and therefore $\delta < \delta_1$. Let us define $M(\gamma, x)$ and $N(\gamma, x)$ as follows,

$$M(\gamma, x) = \left\{\gamma : \min_{\{(s,j):s<j\}} d(\gamma_j, \gamma_s) \geq x\right\}, \quad N(\gamma, x) = \{\gamma : |\gamma_j - \gamma_{0j}| \leq x; j = 1, \ldots, k_0\}$$

Then,

$$\begin{aligned}\pi(||\gamma-\gamma_0||_1\leq\epsilon) &\geq \int_{\{||\gamma-\gamma_0||_1\leq\epsilon\}\cap M(\gamma,\delta)}\pi(\gamma)d\gamma\\ &\gtrsim \int_{\{||\gamma-\gamma_0||_1\leq\epsilon\}\cap M(\gamma,\delta)}\epsilon^{k_0}\prod_{j=1}^{k_0}g_0(\gamma_j)d\gamma\\ &\gtrsim \int_{N(\gamma,\epsilon/k_0)\cap M(\gamma,\delta_1)}\epsilon^{k_0}\prod_{j=1}^{k_0}g_0(\gamma_j)d\gamma\end{aligned}$$

Now let us show that $N(\gamma,\epsilon/k_0)\subseteq M(\gamma,\delta_1)$. Consider the pair (s,j) with $s\neq j$. Without loss of generality assume $\gamma_{0s}>\gamma_{0j}$. Now, consider the possible values of (γ_j,γ_s) contained in the set $N(\gamma,\epsilon/k_0)$. The smallest distance between values of γ_s and γ_j contained in $N(\gamma,\epsilon/k_0)$ is

$$(\gamma_{0s}-\epsilon/k_0)-(\gamma_{0j}+\epsilon/k_0)\geq\epsilon_1-2\epsilon/k_0\geq\epsilon_1(1-1/k_0)=\delta_1$$

Since the previous holds for any pair (s,j) we have $N(\gamma,\epsilon/k_0)\subseteq M(\gamma,\delta_1)$. Therefore,

$$\begin{aligned}\pi(||\gamma-\gamma_0||_1\leq\epsilon) &\gtrsim \int_{N(\gamma,\epsilon/k_0)}\epsilon^{k_0}\prod_{j=1}^{k_0}g_0(\gamma_j)d\gamma\\ &\gtrsim \epsilon^{k_0}\exp\{-g_1k_0\log(1/\epsilon)\}\\ &\gtrsim \exp\{-(g_1+1)k_0\log(1/\epsilon)\}\end{aligned}$$

for a constant $g_1>0$. □

Proof of theorem 1. Only for this proof and for ease of notation the density f will be referred as f_θ. Define the non identifiability set as $T=\{\theta:f_\theta=f_0\}$. In order to define each vector in T, let $0=t_0<t_1<t_2\ldots<t_{k_0}\leq k$ and $\gamma_j=\gamma_{0i}$ for $j\in I_i=\{t_{i-1}+1,t_i\}$. Let $p_{0i}=\sum_{j=t_{i-1}+1}^{t_i}p_j$ and $p_j=0$ for $j>t_{k_0}$. Define $q_j=p_j/p_{0i}$ for $j\in I_i$. Define $A_n=\left\{\min_{\sigma\in S_k}\left(\sum_{i=1}^{k-k_0}p_{\sigma(i)}\right)>\delta_nM_n\right\}$ and $A_n'=A_n\cap\{\|f-f_0\|_1\leq M\delta_n\}$. Let $D_n=\int_{\{\|f-f_0\|_1<\delta_n\}}\exp(l_n(\theta)-l_n(\theta_0))d(\pi\times\lambda)(\theta)$ with $l_n(\theta_0)$ being the log-likelihood evaluated at θ_0. Along the line of [5]'s proof, to prove theorem 1 we need to show that for any $\epsilon>0$ there are positive constants m_1,m_2 and a permutation $\sigma\in S_k$ such that

$$D_n\geq m_1n^{-s(k_0,\alpha)/2} \tag{4}$$

$$\Pi(A_n')\leq m_2\delta_n^{s(k_0,\alpha)}M_n^{\bar{\alpha}-m/2-r_2} \tag{5}$$

with $s(k_0,\alpha)=k_0-1+mk_0+\sum_{j=1}^{k-k_0}\alpha_{\sigma(j)}$. Following [5]'s proof, we can show that, under condition B5, (4) is satisfied for sufficiently large n. Concerning (5), [5] showed that on A_n', there is a set I_i containing indices j_1 and j_2 such that

$$|\gamma_{j_1}-\gamma_{0i}|\leq(\delta_n/q_{j_1})^{1/2},\quad |\gamma_{j_2}-\gamma_{0i}|\leq(\delta_n/q_{j_2})^{1/2}$$

with $q_{j_1}>\epsilon/k_0$ and $q_{j_2}>\delta_nM_n/2$. The triangle inequality implies

$$|\gamma_{j_1}-\gamma_{j_2}|\leq 2\left\{\delta_n/\min(q_{j_1},q_{j_2})\right\}^{1/2}$$

Now, for sufficiently large n, $\min(q_{j_1},q_{j_2})>\delta_nM_n/2$ and therefore $|\gamma_{j_1}-\gamma_{j_2}|\lesssim M_n^{-1/2}$. Since g is bounded above by a positive constant, it exists a constant $c>0$ such that

$$h(\gamma)\leq cg\left\{d(\gamma_{j_1},\gamma_{j_2})\right\}\leq cg\left(M_n^{-1/2}\right) \tag{6}$$

for $\gamma\in A'$. Let the prior probability of the set A_n' be defined as $\Pi(A_n')=\int_{A_n'}d(\pi\times\lambda)(\gamma\times p)$. To find an upper bound for this integral, directly apply the proof of [5] showing that $\Pi(A_n')\leq g\left(M_n^{-1/2}\right)\delta_n^{s(k_0,\alpha)}M_n^{\bar{\alpha}-m/2}$. By assumption, for sufficiently large n, $g\left(M_n^{-1/2}\right)\leq r_1M_n^{-r_2}$. Letting $s_{r_2}=r_2+m/2-\bar{\alpha}$, it follows

$$\Pi(A_n')\leq M_n^{-s_{r_2}}\delta_n^{s(k_0,\alpha)}$$

□

References

[1] Paul Damien, Jon Wakefield, and Stephen Walker. Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variables. *Journal of the Royal Statistical Society B*, 61:331–344, 1999.

[2] Subhashis Ghosal, J. K. Ghosh, and Aad W. Van Der Vaart. Convergence Rates of Posterior Distributions. *The Annals of Statistics*, 28:500–531, 2000.

[3] M. Kalli, J. E. Griffin, and S. G. Walker. Slice Sampling Mixture Models. *Statistics and Computing*, 21:93–105, 2011.

[4] R. M. Neal. Slice Sampling. *The Annals of Statistics*, 31:705–767, 2003.

[5] J. Rousseau and K. Mengersen. Asymptotic Behaviour of the Posterior Distribution in Over-Fitted Models. *Journal of the Royal Statistical Society B*, 73:689–710, 2011.

[6] Catia Scricciolo. Posterior Rates of Convergence for Dirichlet Mixtures of Exponential Power Densities. *Electronic Journal of Statistics*, 5:270–308, 2011.

Table 1: Percentiles 2·5th, 50th and 97·5th of location parameters involved in the two components with highest weights and sum of extra weights under repulsive atoms and non-repulsive atoms for different values of $\tilde{\alpha}$ under dataset (IIb)

	n=100				n=1000			
	Non-Repulsive			Repulsive	Non-Repulsive			Repulsive
$\tilde{\alpha}$	$1/3$	$1/10$	$1/100$	$1/3$	$1/3$	$1/10$	$1/100$	$1/3$
Parameter 1 (truth 1·50)								
2·5	1·59	1·61	1·61	1·53	1·51	1·52	1·52	1·49
50	1·62	1·61	1·61	1·64	1·52	1·52	1·52	1·53
97·5	1·68	1·64	1·63	1·75	1·54	1·53	1·52	1·56
Parameter 2 (truth -1·50)								
2·5	-1·50	-1·48	-1·42	-1·70	-1·51	-1·50	-1·49	-1·55
50	-1·39	-1·39	-1·39	-1·52	-1·48	-1·48	-1·48	-1·49
97·5	-1·21	-1·28	-1·32	-1·34	-1·44	-1·47	-1·47	-1·44
Extra weights sum ($\times 10^3$)								
2·5	4	0	0	1	1	0	0	0
50	69	8	1	10	12	1	1	1
97·5	277	139	15	43	138	25	4	4

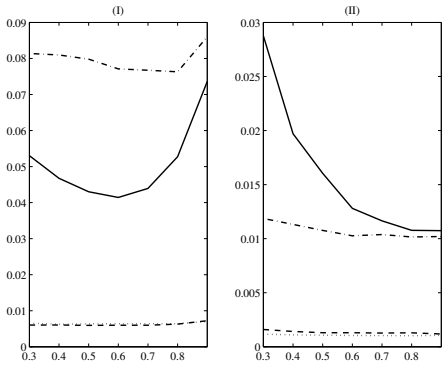

Figure 1: Plot of KL divergence (I) and sum of extra weights (II) for different choice of separation probability (x axis) under dataset (IIa) for $n = 100$ (solid) and $n = 1000$ (dash) and dataset (IIb) for $n = 100$ (dash-dot) and $n = 1000$ (dot) under a mixture of six components

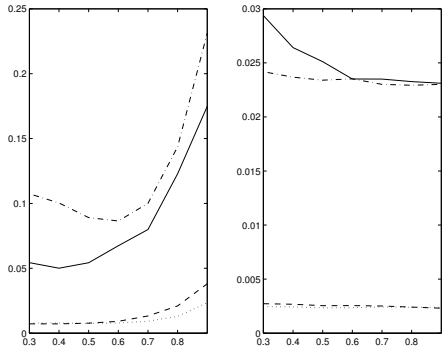

Figure 2: Plot of KL divergence (I) and sum of extra weights (II) for different choice of separation probability (x axis) under dataset (IIa) for $n = 100$ (solid) and $n = 1000$ (dash) and dataset (IIb) for $n = 100$ (dash-dot) and $n = 1000$ (dot) under a mixture of ten components