Supplementary Material of the paper: A Data-Driven Allocation Sampler for Making Inference on the Number of Clusters in the Dirichlet allocation Model via Bayesian Mixture Modelling

This supplementary material (SM) presents the acceptance probability for split-merge movements and some additional results obtained with application of the proposed SDAS algorithm to the simulated datasets.

Appendix 1: Acceptance Probability

From equation (14) of the paper, the acceptance probability for a split movement is $\Psi[\mathbf{\Phi}^{sp}|\mathbf{\Phi}] = min(1, A^{sp})$, where

$$A^{sp} = \frac{P(\mathbf{y}|\mathbf{c}^{sp}, \boldsymbol{\theta}_{k_{\mathbf{c}}^{sp}}, k_{\mathbf{c}^{sp}})}{P(\mathbf{y}|\mathbf{c}, \boldsymbol{\theta}_{k_{\mathbf{c}}}, k_{\mathbf{c}})} \frac{\pi(\mathbf{c}^{sp}|\gamma)}{\pi(\mathbf{c}|\gamma)} \frac{\pi(\boldsymbol{\theta}^{sp}|\eta, k_{\mathbf{c}^{sp}})}{\pi(\boldsymbol{\theta}^{sp}|\eta, k_{\mathbf{c}})} \frac{\prod_{j=k_{\mathbf{c}^{sp}}+1}^{n} \pi(\theta_j|\eta_j)}{\prod_{j=k_{\mathbf{c}}+1}^{n} \pi(\theta_j|\eta_j)} \frac{q[\mathbf{\Phi}|\mathbf{\Phi}^{sp}]}{q[\mathbf{\Phi}^{sp}|\mathbf{\Phi}]},$$

The ratio of the joint probabilities of $\mathbf{y}$ conditional on latent indicator variables and component parameters is given by

$$\frac{P(\mathbf{y}|\mathbf{c}^{sp}, \boldsymbol{\theta}_{k_{\mathbf{c}}^{sp}}, k_{\mathbf{c}^{sp}})}{P(\mathbf{y}|\mathbf{c}, \boldsymbol{\theta}_{k_{\mathbf{c}}}, k_{\mathbf{c}})} = \frac{L\left(\theta_{j_1}^{sp}|D_{j_1}\right) L\left(\theta_{j_2}^{sp}|D_{j_2}\right)}{L\left(\theta_j|D_j\right)}, \tag{20}$$

where $L(\theta_m|D_m) = \prod_{D_m} f(y_i|\theta_m)$ is the likelihood function for component m, for $m = \{j, j_1, j_2\}$.

From Equation (7) of the paper, the ratio of the joint prior probability for latent indicator variables is

$$\frac{\pi\left(\mathbf{c}^{sp}|\gamma, k_{\mathbf{c}}^{sp}\right)}{\pi\left(\mathbf{c}|\gamma, k_{\mathbf{c}}\right)} = \frac{\Gamma(n_{j_1})\Gamma(n_{j_2})}{\Gamma(n_j)}. \tag{21}$$

The prior distributions ratio for component parameters is

$$\frac{\pi\left(\boldsymbol{\theta}_n|\boldsymbol{\eta}, k_{\mathbf{c}}^{sp}\right)}{\pi\left(\boldsymbol{\theta}_n|\boldsymbol{\eta}, k_{\mathbf{c}}\right)} = \frac{\prod_{j=1}^{n} \pi_G\left(\theta_j^{sp}|\eta_j\right)}{\prod_{j=1}^{n} \pi_G\left(\theta_j|\eta_j\right)} = \frac{\pi_G\left(\theta_{j_1}^{sp}|\eta_{j_1}\right) \pi_G\left(\theta_{j_2}^{sp}|\eta_{j_2}\right)}{\pi_G\left(\theta_j|\eta_j\right) \pi_G\left(\theta_n|\eta_n\right)}. \tag{22}$$

From Equation (17) of the paper, the transition probability ratio for the split proposal is

$$\frac{q[\mathbf{\Phi}|\mathbf{\Phi}^{sp}]}{q[\mathbf{\Phi}^{sp}|\mathbf{\Phi}]} = \frac{P_{me|k_{\mathbf{c}}^{sp}}}{P_{sp|k_{\mathbf{c}}}} \frac{P_{j_1,j_2|k_{\mathbf{c}}^{sp}}}{P_{j|\kappa_2}} \frac{1}{P_{alloc}^{sp}} \frac{\pi(\theta_j|D_j,\eta_j)\pi(\theta_n|\eta_n)}{\pi(\theta_{j_1}^{sp}|D_{j_1},\eta_{j_1})\pi(\theta_{j_2}^{sp}|D_{j_2},\eta_{j_2})}, \tag{23}$$

where

$$\pi(\theta_m|D_m,\eta_m) = \frac{L(\theta_m|D_m)\pi_G(\theta_m|\eta_m)}{\int L(\theta_m|D_m)\pi_G(\theta_m|\eta_m)d\theta_m}$$

is the posterior distribution for component parameter θ_m, for $m = \{j, j_1, j_2\}$.

Multiplying (20), (21), (22) and (23), the acceptance probability for a split proposal is $\Psi[\mathbf{\Phi}^{sp}|\mathbf{\Phi}] = min(1, A^{sp})$ for

$$A^{sp} = \frac{\mathbf{I}(D_{j_1})\mathbf{I}(D_{j_2})}{\mathbf{I}(D_j)} \frac{\Gamma(n_{j_1})\Gamma(n_{j_2})}{\Gamma(n_j)} \frac{Q^{sp}}{P_{alloc}},$$

where

$$Q^{sp} = \frac{P_{me|k_{\mathbf{c}}^{sp}}}{P_{sp|k_{\mathbf{c}}}} \frac{P_{j_1,j_2}}{P_{j|\mathbb{C}_2}} = \begin{cases} \frac{1}{2} & \text{, if } k_{\mathbf{c}} = 1; \\ \left(\frac{1}{2}\right)^{1-\mathbb{I}_{k_{\mathbf{c}}^{sp}}(n)} \frac{3\mathbb{C}_2}{k_{\mathbf{c}}+1} & \text{, if } k_{\mathbf{c}} \in \mathbb{K}_1; \\ 2^{\mathbb{I}_{k_{\mathbf{c}}^{sp}}(n)} \frac{\mathbb{C}_2}{k_{\mathbf{c}}+1} & \text{, if } k_{\mathbf{c}} \in \mathbb{K}_2; \end{cases}$$

with $\mathbb{I}_{k_{\mathbf{c}}^{sp}}(n)$ being an indicator function and the sets $\mathbb{K}_1 = \{2 \leq k_{\mathbf{c}} \leq k-1;$ and $j_1 = 1$ or $j_2 = k_{\mathbf{c}}\}$, $\mathbb{K}_2 = \{2 \leq k_{\mathbf{c}} \leq k-1$ and $j_1 \neq 1$ or $j_2 \neq k_{\mathbf{c}}\}$.

Similarly, the acceptance probability for a merge is $\Psi[\mathbf{\Phi}^{me}|\mathbf{\Phi}] = min(1, A^{me}) = \frac{1}{A^{sp}}$, but with some obvious differences in the substitutions.

Appendix 2: Inference on c and $\boldsymbol{\theta}_{k_{\mathbf{c}}}$

In this Appendix we describe the procedure used to define a configuration for $\mathbf{c}$ and estimate parameters $\boldsymbol{\theta}_{\mathbf{c}}$ conditional on estimate $\tilde{k}_{\mathbf{c}}$ for the number of clusters.

Thus, consider $\tilde{k}_{\mathbf{c}}$ be the estimate for the number of clusters. Following [23], let

(i) $L_{\tilde{k}_{\mathbf{c}}} = \sum_{l=B+1}^{L} \mathbb{I}_{k_{\mathbf{c}}^{(l)}}\left(\tilde{k}_{\mathbf{c}}\right)$ be the number of iterations for which $k_{\mathbf{c}} = \tilde{k}_{\mathbf{c}}$ in $L-B$ iterations, where $\mathbb{I}_{k_{\mathbf{c}}^{(l)}} = 1$ if in l-iteration $k_{\mathbf{c}}^{(l)} = \tilde{k}_{\mathbf{c}}$ and $\mathbb{I}_{k_{\mathbf{c}}^{(l)}} = 0$ otherwise;

(ii) $N_{ij} = \sum_{l=B+1}^{L} \mathbb{I}_{c_i^{(l)}}(j)\mathbb{I}_{k_{\mathbf{c}}^{(l)}}$ be the number of times that observation y_i is allocated in component j in $L_{\tilde{k}_{\mathbf{c}}}$ iterations, where $\mathbb{I}_{c_i^{(l)}}(j) = 1$ if in l-th iteration $c_i = j$ and $\mathbb{I}_{c_i^{(l)}}(j) = 0$ otherwise, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

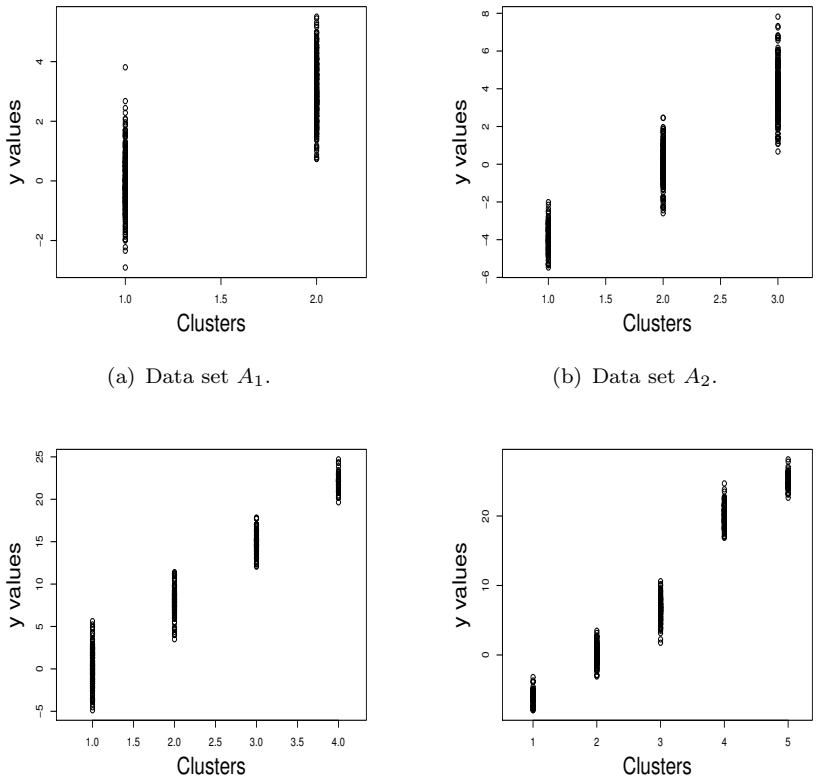
Let $\tilde{P}(c_i = j) = \frac{N_{ij}}{L_{\tilde{k}_{\mathbf{c}}}}$ be the posterior probability that the observation y_i belongs to cluster j, for $j = 1, \ldots, \tilde{k}_{\mathbf{c}}$. If $\tilde{P}(c_i = j) = \max_{1 \leq j' \leq \tilde{k}_{\mathbf{c}}} \left(\tilde{P}(c_i = j')\right)$, then we consider that y_i belongs to cluster j, for $i = 1, \ldots, n$ and $j = 1, \ldots, k_{\mathbf{c}}$.

We estimate parameters θ_j, $j = 1, \ldots, \tilde{k}_{\mathbf{c}}$, considering the average of the generated values, *i.e.*,

$$\tilde{\theta}_j|\tilde{k}_{\mathbf{c}} = \frac{1}{L_{\tilde{k}_{\mathbf{c}}}} \sum_{l=B+1}^{L} \theta_j^{(l)} \mathbb{I}_{k_{\mathbf{c}}^{(l)}} \left(\tilde{k}_{\mathbf{c}}\right).$$

Appendix 3: Generated values

Figure 8 shows the values generated by cluster for datasets A_1 to A_4 for $n = 500$.



(a) Data set A_1.

(b) Data set A_2.

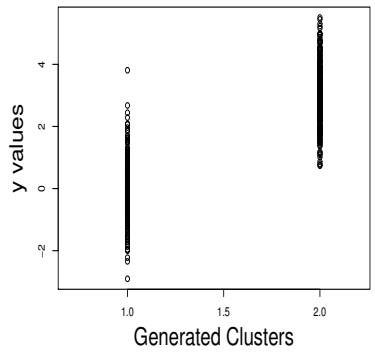
(c) Data set A_3.

(d) Data set A_4.

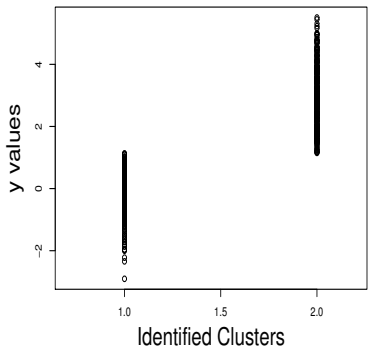
Figure 8. Values generated by cluster for datasets A_1 to A_4 with $n = 500$.

Appendix 4: Results from Artificial data sets

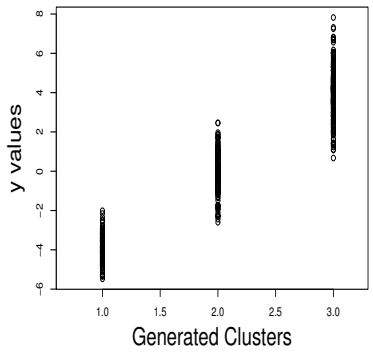
In this section, we present some additional results from simulation study. Figure 9 shows the generated values and the identified clusters by the SDAS algorithm for datasets A_1 to A_4. As one can note, the cluster were satisfactorily identified.



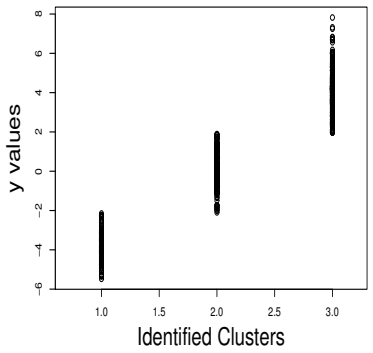
(a) Data set A_1.



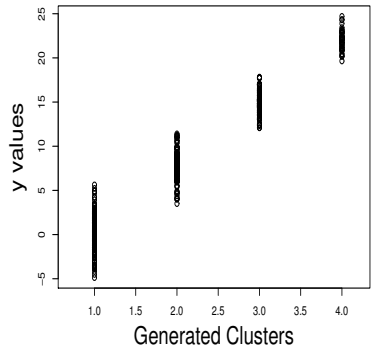
(b) Data set A_1.



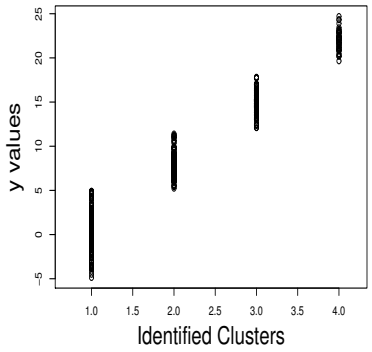
(c) Data set A_2.



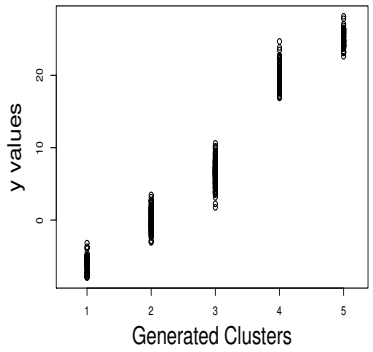
(d) Data set A_2.



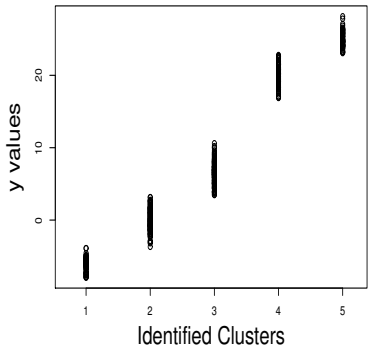
(e) Data set A_3.



(f) Data set A_3.



(g) Data set A_4.



(h) Data set A_4.

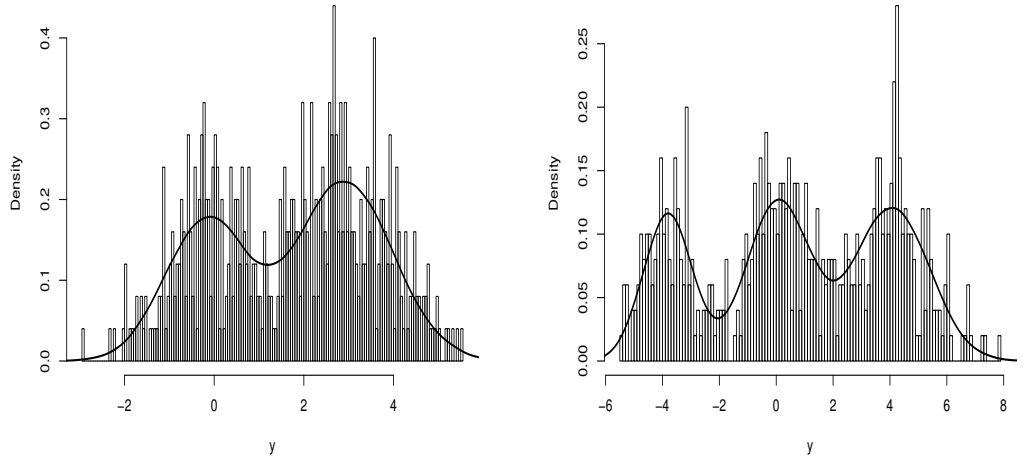
Figure 9. Generated values by cluster and identified clusters by SDAS algorithm, datasets A_1 to A_4.

Table 6 shows the estimates for component parameters of the identified clusters. Estimates for parameters were obtained according to Equation described in page 14, lines 300-302.

Table 6. Estimates for component parameters of the $\tilde{k}_{\mathbf{c}}$ clusters.

Parameter	Data set			
	A_1	A_2	A_3	A_4
μ_1	-0.1884 $(-0.4040, 0.0498)$	-3.8162 $(-4.0087, -3.6235)$	0.3656 $(-0.0066, 0.7621)$	-6.1188 $(-6.3457, -5.8726)$
μ_2	2.9146 $(2.6994, 3.1167)$	0.0951 $(-0.1558, 0.3615)$	8.0019 $(7.6710, 8.3148)$	0.0335 $(-0.3356, 0.4308)$
μ_3	–	4.0610 $(3.7350, 4.3675)$	14.9097 $(14.5892, 15.1972)$	6.5935 $(6.2640, 6.9024)$
μ_4	–	–	22.0472 $(21.7556, 22.3370)$	20.0996 $(19.7348, 20.5063)$
μ_5	–	–	–	24.9200 $(24.4953, 25.2850)$
σ_1^2	0.8015 $(0.5614, 1.1131)$	0.6664 $(0.4586, 0.9363)$	4.9954 $(3.7895, 6.4985)$	0.8813 $(0.6056, 1.2785)$
σ_2^2	1.1406 $(0.8754, 1.5045)$	1.1427 $(0.6555, 1.9545)$	2.2255 $(1.4298, 3.3258)$	2.6316 $(1.6968, 4.0146)$
σ_3^2	–	1.6561 $(1.1567, 2.3125)$	1.9841 $(1.4088, 2.8251)$	$a2.4508$ $(1.7900, 3.3486)$
σ_4^2	–	–	1.1314 $(0.7625, 1/6702)$	2.2023 $(1.5074, 3.2496)$
σ_5^2	–	–	–	1.4701 $(0.9190, 2.3820)$

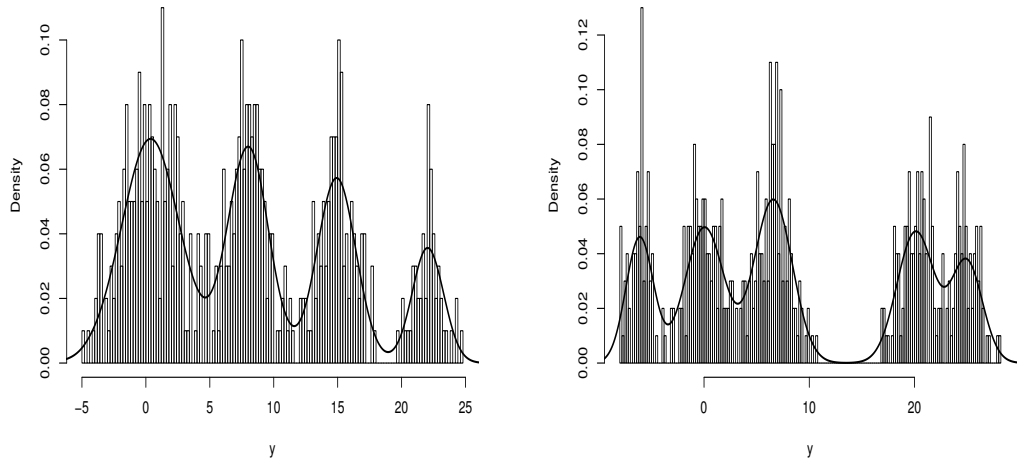
Figures 10 and 11 show the histogram of the observed data and the estimated density function. As one can note the results are satisfactory.



(a) Data set A_1.

(b) Data set A_2.

Figure 10. Histogram of observed data and Estimated density function for datasets A_1 and A_2.



(a) Data set A_3.

(b) Data set A_4.

Figure 11. Histogram of observed data and Estimated density function for datasets A_3 and A_4.

Appendix 4: SDAS codes for Galaxy Dataset

In this Appendix, we present the R codes used in the application of the SDAS algorithm to the Galaxy dataset. The codes in format .txt can be obtained by emailing the first author.

$rm(list = ls(all = TRUE))$
$set.seed(19)$
$library(compiler)$
$enableJIT(3)$
$library(MCMCpack)$
$L = 5000$
$B = 5000$
$salto = 6$
$tx.acs = 0$
$tx.acm = 0$
$km = 10$
$y = c(9172, 9350, 9483, 9558, 9775, 10227, 10406, 16084, 16170, 18419, 18552, 18600, 18927, 19052, 19070, 1$
$n = length(y)$
$y = y/1000$
$gama = 0.1$
$mu0 = (min(y) + max(y))/2$
$lamb = 0.01$
$halfa = 1$
$R = max(y) - min(y)$
$hbeta = halfa/(R^2)$
ID = function(dx, nx){
$DA1 = -(nx/2) * log(2 * hbeta * pi)$
$DA2 = 0.5 * log(lamb) - 0.5 * log(nx + lamb)$

$DA3 = lgamma(halfa + ((nx + 1)/2)) - lgamma(halfa)$
$DA4 = -(halfa + (nx/2)) * log(1 + ((sum(dx^2) + (lamb * (mu0^2)))/(2 * hbeta)) - (((sum(dx) + (lamb * mu0))^2)/(2 * hbeta * (nx + lamb))))$
$(DA1 + DA2 + DA3 + DA4)$ }
$KL = function(dx1, dx)\{$
$nx1 = length(dx1)$
$nx = length(dx)$
$alfa1e = halfa + ((nx1 + 1)/2)$
$alfa0e = halfa + ((nx + 1)/2)$
$beta1e = hbeta + ((sum(dx1^2) + (lamb * (mu0^2)))/2) - (((sum(dx1) + (lamb * mu0))^2)/(2 * (nx1 + lamb)))$
$beta0e = hbeta + ((sum(dx^2) + (lamb * (mu0^2)))/2) - (((sum(dx) + (lamb * mu0))^2)/(2 * (nx + lamb)))$
$mi1e = (sum(dx1) + (lamb * mu0))/(nx1 + lamb)$
$mi0e = (sum(dx) + (lamb * mu0))/(nx + lamb)$
$KL1 = 0.5 * log(nx1 + lamb) - 0.5 * log(nx + lamb) - 0.5 + ((nx + lamb)/(2 * (nx1 + lamb)))$
$KL2 = (((nx + lamb) * ((mi1e - mi0e)^2))/2) * (alfa1e/beta1e)$
$KL3 = alfa0e * log(beta1e) - alfa0e * log(beta0e)$
$KL4 = lgamma(alfa0e) - lgamma(alfa1e)$
$KL5 = (alfa1e - alfa0e) * digamma(alfa1e)$
$KL6 = -(beta1e - beta0e) * (alfa1e/beta1e)$
$(KL1 + KL2 + KL3 + KL4 + KL5 + KL6)$ }
$mu = rep(1000, km)$ $sig = rep(1000, km)$
$kc = numeric()$
$z = matrix(0, nrow = n, ncol = km + 1)$
$ind.y = matrix(0, nrow = n, ncol = km)$
$nj = rep(0, km)$
$syj = rep(0, km)$
$syj2 = rep(0, km)$
$prop = matrix(0, nrow = n, ncol = km)$
$kc[1] = 1$
$z[, 1] = 1$
$ind.y[, 1] = y$
$nj[1] = n$
$syj[1] = sum(y)$
$syj2[1] = sum(y^2)$
$mu[1] = mean(y)$
$sig[1] = var(y)$
$kc.est = numeric()$
$kc.est[1] = 1$
$l0 = B + 1$
$l1 = 1$
$for(lin2 : L)\{$
$zt = z$
$for(iin1 : n)\{$
$p = rep(0, km)$
$for(jin1 : kc)\{$
$if(z[i, j] == 1 \& nj[j] > 1)$
$p[j] = ((nj[j] - 1)/(n + gama - 1)) * dnorm(y[i], mu[j], sqrt(sig[j]))$

$if(z[i,j] == 0 \& nj[j] > 0)$
$p[j] = (nj[j]/(n + gama - 1)) * dnorm(y[i], mu[j], sqrt(sig[j]))$ }
$p0 = (gama/(n + gama - 1)) * exp(ID(y[i], 1))$
$b = 1/(sum(p) + p0)$
$Pr = c(b * p, b * p0)$
$z[i,] = rmultinom(1, 1, Pr)$
$if(z[i,(km + 1)] == 0)$
$ind.y[i,] = z[i, 1 : km] * y[i]$ }
$if(z[i,(km + 1)] == 1)$ {
$kc = kc + 1$
$z[i,] = 0$
$z[i, kc] = 1$
$ind.y[i,] = 0$
$ind.y[i, kc] = y[i]$
$alfaj = halfa + 1$
$betaj = hbeta + (((y[i]^2) + (lamb * (mu0^2)))/2) - ((y[i] + lamb * mu0)^2)/(2 * (1 + lamb))$
$sig[kc] = 1/rgamma(1, alfaj, betaj)$
$muj = (y[i] + lamb * mu0)/(1 + lamb)$
$sigj = sig[kc]/(1 + lamb)$
$mu[kc] = rnorm(1, muj, sqrt(sigj))$ }
$for(jin1 : kc)\{$
$nj[j] = sum(z[, j])$ } }
$kc = 0$
$for(jin1 : km)\{$
$nj[j] = sum(z[, j])$
$if(nj[j] > 0)$
$kc = kc + 1$
$if(nj[j] == 0)$ {
$mu[j] = 1000$
$sig[j] = 1000$ } }
$mu.ord = sort(mu)$
$z.ord = z$
$sig.ord = sig$
$indy.ord = ind.y$
$for(gin1 : kc)\{$
$for(win1 : km)\{$
$if(mu.ord[g] == mu[w])$ {
$sig.ord[g] = sig[w]$
$z.ord[, g] = z[, w]$
$indy.ord[, g] = ind.y[, w]$ } } }
$mu = mu.ord$
$sig = sig.ord$
$z = z.ord$
$ind.y = indy.ord$
$mu[(kc + 1) : km] = 1000$
$sig[(kc + 1) : km] = 1000$
$for(gin1 : kc)\{$
$nj[g] = sum(z[, g])$
$syj[g] = sum(ind.y[, g])$
$syj2[g] = sum(ind.y[, g]^2)$ }

$sp < -0.5$
$me < -0.5$
$if(kc == 1)$ {
$sp = 1$
$mg = 0$ }
$ind.mov = rbinom(1, 1, sp)$
$if(ind.mov == 1)$ {
$if(kc == 1)$ {
$k1 = 1$
$k2 = 2$ }
$if(kc > 1)$ {
$k1 = sample(seq(1, kc, 1), 1)$ $k2 = k1 + 1$ }
$if(nj[k1] > 1)$ {
$kc = kc + 1$
$ind.ysp = ind.y$
$i1 = k2 + 1$
$i2 = km$
$ind.ysp[, i1 : i2] = ind.y[, k2 : (km - 1)]$
$ind.ysp[, k1 : k2] = 0$
$z.sp = z$
$z.sp[, i1 : i2] = z[, k2 : (km - 1)]$
$z.sp[, k1 : k2] = 0$
$y.sp = ind.y[, k1]$
$y.sp = y.sp[y.sp! = 0]$
$isp = match(y.sp, ind.y[, k1])$
$ns = length(y.sp)$
$Palloc < -1$
$if(ns > 2)$ {
$cl = rep(1, ns)$
$ymin = min(y.sp)$
$ymax = max(y.sp)$
$Dj1 = ymin$
$Dj2 = ymax$
$nj1 = 1$
$nj2 = 1$
$Pm = rep(1, ns)$
$for(iin1 : ns)${
$if(y.sp[i] == ymax)$
$cl[i] < -2$
$if(y.sp[i]! = ymin\&y.sp[i]! = ymax)$ {
$xj1 = c(Dj1, y.sp[i])$ $xj2 = c(Dj2, y.sp[i])$ $pj1 = KL(xj1, Dj1)$ $pj2 = KL(xj2, Dj2)$ $P1 = pj2/(pj1 + pj2)$
$P2 = 1 - P1$
$Pm[i] = P1$
$Im = rbinom(1, 1, P1)$
$if(Im == 1)$ {
$Dj1 = c(Dj1, y.sp[i])$
$nj1 < -nj1 + 1$ }
$if(Im == 0)$ {
$Dj2 = c(Dj2, y.sp[i])$

$cl[i] = 2$
$Pm[i] = P2$
$nj2 = nj2 + 1$ } } }
$Palloc < -prod(Pm)$ }
$if(ns == 2)$
$cl = c(1, 2)$
$for(j1in1 : ns)\{$
$if(cl[j1] == 1)$ {
$ind.ysp[isp[j1], k1] = y[isp[j1]]$
$z.sp[isp[j1], k1] = 1$ }
$if(cl[j1] == 2)$ {
$ind.ysp[isp[j1], k2] = y[isp[j1]]$
$z.sp[isp[j1], k2] = 1$ } }
$nj.sp = nj$
$syj.sp = syj$
$syj2.sp = syj2$
$for(gin1 : kc)\{$
$nj.sp[g] = sum(z.sp[, g])$
$syj.sp[g] = sum(ind.ysp[, g])$
$syj2.sp[g] = sum(ind.ysp[, g]^2)$ }
$mu.sp = mu$
$sig.sp = sig$
$mu.sp[i1 : i2] = mu[k2 : (km - 1)]$
$sig.sp[i1 : i2] = sig[k2 : (km - 1)]$
$mu.sp[k1 : k2] = 0$
$sig.sp[k1 : k2] = 0$
$alfaj = halfa + ((nj.sp[k1] + 1)/2)$
$betaj = hbeta + ((syj2.sp[k1] + (lamb * (mu0^2)))/2) - (((syj.sp[k1] + (lamb * mu0))^2)/(2 * (nj.sp[k1] + lamb)))$
$sig.sp[k1] = 1/rgamma(1, alfaj, betaj)$
$alfaj = halfa + ((nj.sp[k2] + 1)/2)$
$betaj = hbeta + ((syj2.sp[k2] + (lamb * (mu0^2)))/2) - (((syj.sp[k2] + (lamb * mu0))^2)/(2 * (nj.sp[k2] + lamb)))$
$sig.sp[k2] = 1/rgamma(1, alfaj, betaj)$
$muj = (syj.sp[k1] + lamb * mu0)/(nj.sp[k1] + lamb)$
$sigj = sig.sp[k1]/(nj.sp[k1] + lamb)$
$mu.sp[k1] = rnorm(1, muj, sqrt(sigj))$
$muj = (syj.sp[k2] + lamb * mu0)/(nj.sp[k2] + lamb)$
$sigj = sig.sp[k2]/(nj.sp[k2] + lamb)$
$mu.sp[k2] = rnorm(1, muj, sqrt(sigj))$
$if(mu.sp[k1] > mu.sp[k2])$ {
$zz1 = z.sp[, k2]$
$zz2 = z.sp[, k1]$
$iy1 = ind.ysp[, k2]$
$iy2 = ind.ysp[, k1]$
$m1 = mu.sp[k2]$
$m2 = mu.sp[k1]$
$si1 = sig.sp[k2]$
$si2 = sig.sp[k1]$
$nj1 = nj.sp[k2]$

$nj2 = nj.sp[k1]$
$sy1 = syj.sp[k2]$
$sy2 = syj.sp[k1]$
$sy21 = syj2.sp[k2]$
$sy22 = syj2.sp[k1]$
$z.sp[, k1] = zz1$
$z.sp[, k2] = zz2$
$ind.ysp[, k1] = iy1$
$ind.ysp[, k2] = iy2$
$mu.sp[k1] = m1$
$mu.sp[k2] = m2$
$sig.sp[k1] = si1$
$sig.sp[k2] = si2$
$nj.sp[k1] = nj1$
$nj.sp[k2] = nj2$
$syj.sp[k1] = sy1$
$syj.sp[k2] = sy2$
$syj2.sp[k1] = sy21$
$syj2.sp[k2] = sy22$ }
$ind.nad = 0$
$for(jin1 : kc)\{$
$if(mu.sp[j] > mu.sp[(j + 1)])$
$ind.nad = 1$ }
$As = 0$
$if(ind.nad == 0)$ {
$Q.sp = 0.5$
$if((kc - 1) > 1)$ {
$if(k1 == 1)$
$Q.sp = 3 * (kc - 1)/(2 * kc)$
$if(k1! = 1)$
$Q.sp = (kc - 1)/kc$ }
$Ga = lgamma(nj.sp[k1]) + lgamma(nj.sp[k2]) - lgamma(nj[k1])$ $As = Q.sp * (1/Palloc) * exp(ID(ind.ysp[, k1], nj.sp[k1]) + ID(ind.ysp[, k2], nj.sp[k2]) - ID(ind.y[, k1], nj[k1]) + Ga)$ }
$if(As == "NaN")$
$As = 0$
$alpha.s = min(1, As)$
$u.ts = runif(1)$
$if(u.ts < alpha.s)$ {
$z = z.sp$
$ind.y = ind.ysp$
$mumu.sp$
$sig = sig.sp$
$tx.acs = tx.acs + 1$
$nj = nj.sp$
$syj = syj.sp$
$syj2 = syj2.sp$ }
$if(u.ts >= alpha.s)$
$kc = kc - 1$ } }
$if(ind.mov == 0)$ {

$kc = kc - 1$
$k1 < -sample(seq(1, kc, 1), 1)$
$if(k1 == 1)$
$k2 < -2$
$if(k1! = 1)$ {
$co < -rbinom(1, 1, 0.5)$
$if(co == 1)$ {
$k2 = k1$
$k1 = k1 - 1$ }
$if(co == 0)$
$k2 = k1 + 1$ }
$z.me = z$
$z.me[, k1] = z[, k1] + z[, k2]$
$me2km - 1$
$z.me[, (k2 : me2)] = z[, (k2 + 1) : km]$
$ind.yme = ind.y$
$ind.yme[, k1] = ind.y[, k1] + ind.y[, k2]$
$ind.yme[, k2 : me2] = ind.y[, (k2 + 1) : km]$
$nj.me = nj$
$syj.me = syj$
$syj2.me = syj2$
$nj.me[k1] = nj[k1] + nj[k2]$
$syj.me[k1] = syj[k1] + syj[k2]$
$syj2.me[k1] = syj2[k1] + syj2[k2]$
$nj.me[k2 : me2] = nj[(k2 + 1) : km]$
$syj.me[k2 : me2] = syj[(k2 + 1) : km]$
$syj2.me[k2 : me2] = syj2[(k2 + 1) : km]$
$mu.me = mu$
$sig.me = sig$
$mu.me[k2 : me2] = mu[(k2 + 1) : km]$
$sig.me[k2 : me2] = sig[(k2 + 1) : km]$
$alfaj = halfa + ((nj.me[k1] + 1)/2)$
$betaj = hbeta + ((syj2.me[k1] + (lamb * (mu0^2)))/2) - (((syj.me[k1] + (lamb * mu0))^2)/(2 * (nj.me[k1] + lamb)))$
$sig.me[k1] = 1/rgamma(1, alfaj, betaj)$
$muj = (syj.me[k1] + lamb * mu0)/(nj.me[k1] + lamb)$
$sigj = sig.me[k1]/(nj.me[k1] + lamb)$
$mu.me[k1] = rnorm(1, muj, sqrt(sigj))$
$ind.nad < -0$
$for(jin1 : kc)\{$
$if(mu.me[j] > mu.me[(j + 1)])$
$ind.nad < -1$ }
$y.me1 = ind.y[, k1]$
$y.me1 = y.me1[y.me1! = 0]$
$ymin = min(y.me1)$
$y.me2 = ind.y[, k2]$
$y.me2 = y.me2[y.me2! = 0]$
$ymax = max(y.me2)$
$y.me = ind.yme[, k1]$
$y.me = y.me[y.me! = 0]$

$isp = match(y.me, ind.yme[, k1])$
$ns = length(y.me)$
$Palloc = 1$
$cl = rep(1, ns)$
$Dj1 = ymin$
$Dj2 = ymax$
$nj1 = 1$
$nj2 = 1$
$Pm = rep(1, ns)$
$for(iin1 : ns)\{$
$if(y.me[i]! = ymin \& y.me[i]! = ymax)\ \{$
$xj1 = c(Dj1, y.me[i])$
$xj2 = c(Dj2, y.me[i])$
$pj1 = KL(xj1, Dj1)$
$pj2 = KL(xj2, Dj2)$
$P1 = pj2/(pj1 + pj2)$
$P2 = 1 - P1$
$if(z[isp[i], k1] == 1)\ \{$
$Pm[i] < -P1$
$Dj1 < -c(Dj1, y.me[i])$
$nj1 < -nj1 + 1\ \}$
$if(z[isp[i], k2] == 1)\ \{$
$Pm[i] = P2$
$Dj2 = c(Dj2, y.me[i])$
$cl[i] = 2$
$nj2 = nj2 + 1\ \}\ \}\ \}$
$Palloc = prod(Pm)$
$Am = 0$
$if(ind.nad == 0)\ \{$
$if(kc == 1)$
$Q.me = 2$
$if(kc > 1)\ \{$
$if(k1 == 1)$
$Q.me = 2 * (kc + 1)/(3 * kc)$
$if(k1! = 1)$
$Q.me = (kc + 1)/kc\ \}$
$Ga = lgamma(nj.me[k1]) - lgamma(nj[k1]) - lgamma(nj[k2])$
$Am = Q.me * Palloc * exp(ID(ind.yme[, k1], nj.me[k1]) - ID(ind.y[, k1], nj[k1]) - ID(ind.y[, k2], nj[k2]) + Ga)\ \}$
$if(Am == "NaN")$
$Am = 0$
$alpha.m = min(1, Am)$
$u.tm = runif(1)$
$if(u.tm < alpha.m)\ \{$
$z = z.me$
$ind.y = ind.yme$
$mu = mu.me$
$sig = sig.me$
$tx.acm = tx.acm + 1$
$nj = nj.me$

$syj = syj.me$
$syj2 = syj2.me$ }
$if(u.tm > alpha.m)$
$kc = kc + 1$ }
$if(l == l0)\{$
$kc.est[l1] = kc$
$l0 = l0 + salto$
$l1 = l1 + 1$ }
$if(kc == 3)\{$
$prop3 = prop3 + z[, 1:3]$
$mu.est = mu[1:3]$
$sig.est < -sig[1:3]$ }
$mu.prop = mu$
$sig.prop = sig$
$ind.ord = 0$
$for(jin1:kc)\{$
$alfaj = halfa + ((nj[j] + 1)/2)$
$betaj = hbeta + ((syj2[j] + (lamb * (mu0^2)))/2) - (((syj[j] + lamb * mu0)^2)/(2 * (nj[j] + lamb)))$
$sig.prop[j] = 1/rgamma(1, alfaj, betaj)$
$muj = (syj[j] + lamb * mu0)/(nj[j] + lamb)$
$sigj = sig[j]/(nj[j] + lamb)$
$mu.prop[j] = rnorm(1, muj, sqrt(sigj))$
$if(j >= 2)\{$
$if(mu.prop[j] < mu.prop[j-1])$
$ind.ord = 1$ } }
$if(ind.ord == 0)\{$
$mu = mu.prop$
$sig = sig.prop$ } }