Modeling and Inference for Mixtures of Simple Symmetric Exponential Families of p-Dimensional Distributions for Vectors with Binary Coordinates: Supplementary Document

Abhishek Chakraborty and Stephen B. Vardeman

This document supports the findings and analyses in the original research article.

A: Simulation Study on Method Performance (Section 4.1)

The sets of central vectors and variability parameters used to generate the datasets under each setting in the simulation study focusing on method performance are reported below. Also reported are the pixel flip counts between the respective central vectors and the KL divergence values for Figure 1 below. 'KLd-mixture', 'Kld-rf', and 'KLd-uniform' denote the KL divergence measures computed between the "true" model probabilities and the probabilities estimated from the proposed mixture model, the relative frequency distribution, and the uniform distribution respectively.

(i) $N = 500$; $p = 4$; $K_{true} = 2$

(1) $\boldsymbol{\mu}_1 = (1, 1, 1, 1)$, $\boldsymbol{\mu}_2 = (0, 0, 0, 0)$; $\alpha_1 = 0.678$, $\alpha_2 = 0.900$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 4 \\ - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.003	0.014	0.316
2	0.003	0.015	0.316
3	0.001	0.011	0.316
4	0.004	0.014	0.316
5	0.009	0.020	0.316
6	0.015	0.027	0.316
7	0.005	0.017	0.316
8	0.013	0.019	0.316
9	0.006	0.012	0.316
10	0.002	0.009	0.316

(2) $\boldsymbol{\mu}_1 = (1, 0, 0, 1),\ \boldsymbol{\mu}_2 = (1, 1, 0, 1);\ \alpha_1 = 0.344,\ \alpha_2 = 0.807$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 1 \\ - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.005	0.014	0.547
2	0.005	0.012	0.547
3	0.002	0.015	0.547
4	0.002	0.015	0.547
5	0.001	0.010	0.547
6	0.003	0.016	0.547
7	0.012	0.025	0.547
8	0.004	0.013	0.547
9	0.008	0.032	0.547
10	0.005	0.018	0.547

(3) $\boldsymbol{\mu}_1 = (1, 1, 0, 1),\ \boldsymbol{\mu}_2 = (1, 1, 1, 0);\ \alpha_1 = 0.229,\ \alpha_2 = 0.053$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 2 \\ - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.004	0.025	1.380
2	0.002	0.025	1.380
3	0.005	0.042	1.321
4	0.003	0.048	1.407
5	0.002	0.014	1.433
6	0.007	0.021	1.484
7	0.009	0.021	1.434
8	0.007	0.077	1.380
9	0.005	0.102	1.250
10	0.002	0.024	1.380

(4) $\boldsymbol{\mu}_1 = (1, 1, 0, 0),\ \boldsymbol{\mu}_2 = (1, 0, 1, 1);\ \alpha_1 = 0.562,\ \alpha_2 = 0.358$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 3 \\ - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.001	0.009	0.604
2	0.003	0.012	0.604
3	0.001	0.012	0.604
4	0.009	0.015	0.604
5	0.003	0.026	0.604
6	0.002	0.020	0.604
7	0.013	0.025	0.604
8	0.005	0.019	0.604
9	0.010	0.020	0.604
10	0.003	0.015	0.604

(5) $\boldsymbol{\mu}_1 = (1, 1, 0, 1),\ \boldsymbol{\mu}_2 = (0, 1, 0, 0);\ \alpha_1 = 0.914,\ \alpha_2 = 0.054$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 2 \\ - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.007	0.014	0.820
2	0.006	0.014	0.820
3	0.003	0.013	0.820
4	0.003	0.009	0.820
5	0.003	0.016	0.820
6	0.003	0.019	0.820
7	0.006	0.026	0.820
8	0.009	0.017	0.820
9	0.006	0.021	0.820
10	0.001	0.012	0.820

(ii) $N = 1000$; $p = 6$; $K_{true} = 7$

(1) $\boldsymbol{\mu}_1 = (1,0,0,0,0,1)$, $\boldsymbol{\mu}_2 = (1,1,1,1,0,1)$, $\boldsymbol{\mu}_3 = (0,0,1,0,1,1)$, $\boldsymbol{\mu}_4 = (0,0,1,1,0,1)$, $\boldsymbol{\mu}_5 = (0,1,0,0,1,1)$, $\boldsymbol{\mu}_6 = (1,0,0,1,1,1)$, $\boldsymbol{\mu}_7 = (0,0,0,0,0,1)$;

$\alpha_1 = 0.605$, $\alpha_2 = 0.724$, $\alpha_3 = 0.191$, $\alpha_4 = 0.516$, $\alpha_5 = 0.261$, $\alpha_6 = 0.238$, $\alpha_7 = 0.609$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 3 & 3 & 3 & 3 & 2 & 1 \\ - & 0 & 4 & 2 & 4 & 3 & 4 \\ - & - & 0 & 2 & 2 & 3 & 2 \\ - & - & - & 0 & 4 & 3 & 2 \\ - & - & - & - & 0 & 3 & 2 \\ - & - & - & - & - & 0 & 3 \\ - & - & - & - & - & - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.009	0.046	0.787
2	0.012	0.039	0.776
3	0.006	0.091	0.826
4	0.015	0.069	0.804
5	0.012	0.035	0.776
6	0.007	0.030	0.776
7	0.010	0.047	0.787
8	0.007	0.030	0.776
9	0.012	0.049	0.790
10	0.014	0.042	0.776

(2) $\boldsymbol{\mu}_1 = (0,1,0,1,1,0),\ \boldsymbol{\mu}_2 = (1,0,1,1,0,0),\ \boldsymbol{\mu}_3 = (0,0,1,1,0,0),\ \boldsymbol{\mu}_4 = (1,0,0,0,0,1),$
$\boldsymbol{\mu}_5 = (1,0,1,0,1,1),\ \boldsymbol{\mu}_6 = (0,1,0,0,1,1),\ \boldsymbol{\mu}_7 = (0,1,0,1,0,1);$
$\alpha_1 = 0.289,\ \alpha_2 = 0.743,\ \alpha_3 = 0.333,\ \alpha_4 = 0.489,\ \alpha_5 = 0.581,\ \alpha_6 = 0.720,\ \alpha_7 = 0.280$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 4 & 3 & 5 & 5 & 2 & 2 \\ - & 0 & 1 & 3 & 3 & 6 & 4 \\ - & - & 0 & 4 & 4 & 5 & 3 \\ - & - & - & 0 & 2 & 3 & 3 \\ - & - & - & - & 0 & 3 & 5 \\ - & - & - & - & - & 0 & 2 \\ - & - & - & - & - & - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.013	0.038	0.680
2	0.006	0.030	0.680
3	0.006	0.037	0.680
4	0.012	0.045	0.680
5	0.011	0.037	0.680
6	0.009	0.037	0.680
7	0.007	0.030	0.680
8	0.004	0.038	0.680
9	0.008	0.036	0.680
10	0.007	0.030	0.680

(3) $\boldsymbol{\mu}_1 = (1,0,0,1,0,1),\ \boldsymbol{\mu}_2 = (0,0,1,0,0,1),\ \boldsymbol{\mu}_3 = (1,1,1,0,1,1),\ \boldsymbol{\mu}_4 = (1,1,0,0,0,1),$
$\boldsymbol{\mu}_5 = (0,1,1,1,0,1),\ \boldsymbol{\mu}_6 = (0,1,0,0,0,0),\ \boldsymbol{\mu}_7 = (1,0,0,1,1,0);$
$\alpha_1 = 0.296,\ \alpha_2 = 0.950,\ \alpha_3 = 0.471,\ \alpha_4 = 0.612,\ \alpha_5 = 0.127,\ \alpha_6 = 0.394,\ \alpha_7 = 0.026$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 3 & 4 & 2 & 3 & 4 & 2 \\ - & 0 & 3 & 3 & 2 & 3 & 5 \\ - & - & 0 & 2 & 3 & 4 & 4 \\ - & - & - & 0 & 3 & 2 & 4 \\ - & - & - & - & 0 & 3 & 5 \\ - & - & - & - & - & 0 & 4 \\ - & - & - & - & - & - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.013	0.050	0.899
2	0.011	0.032	0.887
3	0.006	0.028	0.887
4	0.008	0.060	0.904
5	0.015	0.038	0.887
6	0.010	0.046	0.899
7	0.011	0.054	0.899
8	0.008	0.032	0.887
9	0.012	0.059	0.899
10	0.009	0.030	0.887

(4) $\boldsymbol{\mu}_1 = (0,1,1,0,1,0),\ \boldsymbol{\mu}_2 = (0,0,0,1,1,0),\ \boldsymbol{\mu}_3 = (1,1,1,1,0,0),\ \boldsymbol{\mu}_4 = (1,0,0,1,1,0),$
$\boldsymbol{\mu}_5 = (1,0,1,0,0,0),\ \boldsymbol{\mu}_6 = (0,0,0,0,1,1),\ \boldsymbol{\mu}_7 = (0,0,0,0,1,0);$
$\alpha_1 = 0.135,\ \alpha_2 = 0.810,\ \alpha_3 = 0.056,\ \alpha_4 = 0.307,\ \alpha_5 = 0.391,\ \alpha_6 = 0.756,\ \alpha_7 = 0.923$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 3 & 3 & 4 & 3 & 3 & 2 \\ - & 0 & 4 & 1 & 4 & 2 & 1 \\ - & - & 0 & 3 & 2 & 6 & 5 \\ - & - & - & 0 & 3 & 3 & 2 \\ - & - & - & - & 0 & 4 & 3 \\ - & - & - & - & - & 0 & 1 \\ - & - & - & - & - & - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.007	0.027	0.811
2	0.011	0.065	0.832
3	0.009	0.067	0.841
4	0.009	0.039	0.811
5	0.013	0.038	0.811
6	0.009	0.037	0.811
7	0.006	0.024	0.811
8	0.007	0.054	0.828
9	0.011	0.034	0.811
10	0.007	0.038	0.811

(5) $\boldsymbol{\mu}_1 = (1,1,1,0,1,0),\ \boldsymbol{\mu}_2 = (0,1,0,0,1,1),\ \boldsymbol{\mu}_3 = (1,1,1,0,0,0),\ \boldsymbol{\mu}_4 = (0,1,0,1,1,1),$
$\boldsymbol{\mu}_5 = (1,0,1,0,0,1),\ \boldsymbol{\mu}_6 = (0,1,0,0,0,1),\ \boldsymbol{\mu}_7 = (1,0,0,0,1,1);$
$\alpha_1 = 0.969,\ \alpha_2 = 0.932,\ \alpha_3 = 0.856,\ \alpha_4 = 0.985,\ \alpha_5 = 0.122,\ \alpha_6 = 0.867,\ \alpha_7 = 0.354$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix} 0 & 3 & 1 & 4 & 3 & 4 & 3 \\ - & 0 & 4 & 1 & 4 & 1 & 2 \\ - & - & 0 & 5 & 2 & 3 & 4 \\ - & - & - & 0 & 5 & 2 & 3 \\ - & - & - & - & 0 & 3 & 2 \\ - & - & - & - & - & 0 & 3 \\ - & - & - & - & - & - & 0 \end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.006	0.029	0.430
2	0.016	0.045	0.430
3	0.009	0.037	0.430
4	0.017	0.044	0.430
5	0.014	0.031	0.430
6	0.013	0.043	0.430
7	0.010	0.032	0.430
8	0.004	0.022	0.430
9	0.007	0.031	0.430
10	0.005	0.034	0.430

(iii) $N = 2000$; $p = 8$; $K_{true} = 10$

(1) $\boldsymbol{\mu}_1 = (1,0,0,1,0,1,0,0)$, $\boldsymbol{\mu}_2 = (1,0,0,0,0,1,1,0)$, $\boldsymbol{\mu}_3 = (0,0,1,1,1,0,0,1)$, $\boldsymbol{\mu}_4 = (0,1,1,1,0,0,0,1)$, $\boldsymbol{\mu}_5 = (0,0,1,1,0,0,0,0)$, $\boldsymbol{\mu}_6 = (0,0,0,0,1,0,0,0)$, $\boldsymbol{\mu}_7 = (0,0,0,1,0,0,1,1)$, $\boldsymbol{\mu}_8 = (1,1,0,1,0,0,0,1)$, $\boldsymbol{\mu}_9 = (1,0,1,1,0,0,1,0)$, $\boldsymbol{\mu}_{10} = (1,1,0,1,0,1,0,0)$;

$\alpha_1 = 0.402$, $\alpha_2 = 0.798$, $\alpha_3 = 0.119$, $\alpha_4 = 0.119$, $\alpha_5 = 0.983$, $\alpha_6 = 0.310$, $\alpha_7 = 0.191$, $\alpha_8 = 0.001$, $\alpha_9 = 0.782$, $\alpha_{10} = 0.773$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix}
0 & 2 & 5 & 5 & 3 & 4 & 4 & 3 & 3 & 1 \\
- & 0 & 7 & 7 & 5 & 4 & 4 & 5 & 3 & 3 \\
- & - & 0 & 2 & 2 & 3 & 3 & 4 & 4 & 6 \\
- & - & - & 0 & 2 & 5 & 3 & 2 & 4 & 4 \\
- & - & - & - & 0 & 3 & 3 & 4 & 2 & 4 \\
- & - & - & - & - & 0 & 4 & 5 & 5 & 5 \\
- & - & - & - & - & - & 0 & 3 & 3 & 5 \\
- & - & - & - & - & - & - & 0 & 4 & 2 \\
- & - & - & - & - & - & - & - & 0 & 4 \\
- & - & - & - & - & - & - & - & - & 0
\end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.010	0.298	1.625
2	0.013	0.346	1.669
3	0.007	0.236	1.577
4	0.007	0.270	1.598
5	0.006	0.285	1.605
6	0.007	0.262	1.581
7	0.008	0.303	1.636
8	0.004	0.353	1.672
9	0.008	0.253	1.579
10	0.005	0.285	1.605

(2) $\boldsymbol{\mu}_1 = (0,1,1,0,1,0,1,0),\ \boldsymbol{\mu}_2 = (0,1,1,0,0,0,0,0),\ \boldsymbol{\mu}_3 = (0,1,0,1,1,0,1,0),$
$\boldsymbol{\mu}_4 = (1,0,1,1,0,0,0,0),\ \boldsymbol{\mu}_5 = (1,1,1,0,1,0,0,0),\ \boldsymbol{\mu}_6 = (1,1,1,0,1,1,0,0),$
$\boldsymbol{\mu}_7 = (1,0,1,1,0,0,0,0),\ \boldsymbol{\mu}_8 = (0,0,1,0,1,1,0,1),\ \boldsymbol{\mu}_9 = (0,0,1,0,1,0,0,1),$
$\boldsymbol{\mu}_{10} = (1,0,0,1,0,1,0,1);$
$\alpha_1 = 0.232,\ \alpha_2 = 0.211,\ \alpha_3 = 0.118,\ \alpha_4 = 0.294,\ \alpha_5 = 0.367,\ \alpha_6 = 0.482,$
$\alpha_7 = 0.199,\ \alpha_8 = 0.452,\ \alpha_9 = 0.708,\ \alpha_{10} = 0.783$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix}
0 & 2 & 2 & 5 & 2 & 3 & 5 & 4 & 3 & 8 \\
- & 0 & 4 & 3 & 2 & 3 & 3 & 4 & 3 & 6 \\
- & - & 0 & 5 & 4 & 5 & 5 & 6 & 5 & 6 \\
- & - & - & 0 & 3 & 4 & 0 & 5 & 4 & 3 \\
- & - & - & - & 0 & 1 & 3 & 4 & 3 & 6 \\
- & - & - & - & - & 0 & 4 & 3 & 4 & 5 \\
- & - & - & - & - & - & 0 & 5 & 4 & 3 \\
- & - & - & - & - & - & - & 0 & 1 & 4 \\
- & - & - & - & - & - & - & - & 0 & 5 \\
- & - & - & - & - & - & - & - & - & 0
\end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.006	0.355	1.877
2	0.009	0.372	1.897
3	0.004	0.307	1.847
4	0.008	0.351	1.903
5	0.007	0.354	1.878
6	0.005	0.270	1.838
7	0.007	0.406	1.914
8	0.006	0.442	1.934
9	0.007	0.298	1.849
10	0.007	0.346	1.874

(3) $\boldsymbol{\mu}_1 = (0,1,1,1,1,0,1,0),\ \boldsymbol{\mu}_2 = (0,1,0,1,0,0,0,1),\ \boldsymbol{\mu}_3 = (1,1,1,0,0,0,1,0),$
$\boldsymbol{\mu}_4 = (1,1,1,1,1,0,1,0),\ \boldsymbol{\mu}_5 = (1,1,1,1,0,0,0,0),\ \boldsymbol{\mu}_6 = (0,0,1,1,1,1,1,1),$
$\boldsymbol{\mu}_7 = (0,0,0,1,1,0,0,1),\ \boldsymbol{\mu}_8 = (0,1,0,1,1,1,1,1),\ \boldsymbol{\mu}_9 = (0,0,0,1,1,0,1,0),$
$\boldsymbol{\mu}_{10} = (0,0,0,1,1,0,1,0);$
$\alpha_1 = 0.565,\ \alpha_2 = 0.471,\ \alpha_3 = 0.990,\ \alpha_4 = 0.656,\ \alpha_5 = 0.012,\ \alpha_6 = 0.975,$
$\alpha_7 = 0.678,\ \alpha_8 = 0.681,\ \alpha_9 = 0.321,\ \alpha_{10} = 0.244$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix}
0 & 4 & 3 & 1 & 3 & 3 & 4 & 3 & 2 & 2 \\
- & 0 & 5 & 5 & 3 & 5 & 2 & 3 & 4 & 4 \\
- & - & 0 & 2 & 2 & 6 & 7 & 6 & 5 & 5 \\
- & - & - & 0 & 2 & 4 & 5 & 4 & 3 & 3 \\
- & - & - & - & 0 & 6 & 5 & 6 & 5 & 5 \\
- & - & - & - & - & 0 & 3 & 2 & 3 & 3 \\
- & - & - & - & - & - & 0 & 3 & 2 & 2 \\
- & - & - & - & - & - & - & 0 & 3 & 3 \\
- & - & - & - & - & - & - & - & 0 & 0 \\
- & - & - & - & - & - & - & - & - & 0
\end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.004	0.237	1.384
2	0.006	0.246	1.379
3	0.005	0.216	1.362
4	0.007	0.258	1.396
5	0.005	0.218	1.356
6	0.007	0.248	1.383
7	0.006	0.216	1.354
8	0.009	0.378	1.511
9	0.009	0.218	1.358
10	0.010	0.242	1.382

(4) $\boldsymbol{\mu}_1 = (1,1,0,1,0,0,0,1),\ \boldsymbol{\mu}_2 = (1,0,0,1,1,0,0,0),\ \boldsymbol{\mu}_3 = (0,1,1,1,1,1,1,0),$
$\boldsymbol{\mu}_4 = (1,0,0,0,1,1,1,1),\ \boldsymbol{\mu}_5 = (1,1,0,1,0,1,1,0),\ \boldsymbol{\mu}_6 = (0,0,1,0,0,0,1,0),$
$\boldsymbol{\mu}_7 = (0,0,1,1,0,0,1,0),\ \boldsymbol{\mu}_8 = (1,0,0,0,1,0,1,0),\ \boldsymbol{\mu}_9 = (1,1,1,1,1,1,1,1),$
$\boldsymbol{\mu}_{10} = (0,1,1,0,0,1,0,1);$
$\alpha_1 = 0.837,\ \alpha_2 = 0.901,\ \alpha_3 = 0.826,\ \alpha_4 = 0.094,\ \alpha_5 = 0.219,\ \alpha_6 = 0.381,$
$\alpha_7 = 0.816,\ \alpha_8 = 0.009,\ \alpha_9 = 0.941,\ \alpha_{10} = 0.237$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix}
0 & 3 & 6 & 5 & 3 & 6 & 5 & 5 & 4 & 4 \\
- & 0 & 5 & 4 & 4 & 5 & 4 & 2 & 5 & 7 \\
- & - & 0 & 5 & 3 & 4 & 3 & 5 & 2 & 4 \\
- & - & - & 0 & 4 & 5 & 6 & 2 & 3 & 5 \\
- & - & - & - & 0 & 5 & 4 & 4 & 3 & 5 \\
- & - & - & - & - & 0 & 1 & 3 & 6 & 4 \\
- & - & - & - & - & - & 0 & 4 & 5 & 5 \\
- & - & - & - & - & - & - & 0 & 5 & 7 \\
- & - & - & - & - & - & - & - & 0 & 4 \\
- & - & - & - & - & - & - & - & - & 0
\end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.007	0.233	1.339
2	0.008	0.237	1.341
3	0.007	0.179	1.297
4	0.011	0.220	1.328
5	0.009	0.260	1.351
6	0.005	0.189	1.298
7	0.009	0.243	1.349
8	0.006	0.268	1.370
9	0.011	0.249	1.350
10	0.006	0.222	1.330

(5) $\boldsymbol{\mu}_1 = (1,0,0,0,1,0,1,1),\ \boldsymbol{\mu}_2 = (0,1,0,1,1,0,1,1),\ \boldsymbol{\mu}_3 = (1,0,0,1,0,1,0,1),$
$\boldsymbol{\mu}_4 = (1,0,1,1,1,1,1,0),\ \boldsymbol{\mu}_5 = (0,1,1,1,0,1,1,0),\ \boldsymbol{\mu}_6 = (0,1,0,0,1,0,1,0),$
$\boldsymbol{\mu}_7 = (1,0,1,1,0,1,0,1),\ \boldsymbol{\mu}_8 = (1,0,0,1,1,1,1,1),\ \boldsymbol{\mu}_9 = (0,0,0,0,0,0,1,0),$
$\boldsymbol{\mu}_{10} = (1,1,0,1,0,1,0,0);$
$\alpha_1 = 0.157,\ \alpha_2 = 0.697,\ \alpha_3 = 0.227,\ \alpha_4 = 0.274,\ \alpha_5 = 0.023,\ \alpha_6 = 0.092,$
$\alpha_7 = 0.999,\ \alpha_8 = 0.146,\ \alpha_9 = 0.562,\ \alpha_{10} = 0.324$

The pixel flip counts between the data generating central vectors:

$$\begin{bmatrix}
0 & 3 & 4 & 4 & 7 & 3 & 5 & 2 & 3 & 6 \\
- & 0 & 5 & 5 & 4 & 2 & 6 & 3 & 4 & 5 \\
- & - & 0 & 4 & 5 & 7 & 1 & 2 & 5 & 2 \\
- & - & - & 0 & 3 & 5 & 3 & 2 & 5 & 4 \\
- & - & - & - & 0 & 4 & 4 & 5 & 4 & 3 \\
- & - & - & - & - & 0 & 8 & 5 & 2 & 5 \\
- & - & - & - & - & - & 0 & 3 & 6 & 3 \\
- & - & - & - & - & - & - & 0 & 5 & 4 \\
- & - & - & - & - & - & - & - & 0 & 5 \\
- & - & - & - & - & - & - & - & - & 0
\end{bmatrix}$$

KL divergence values:

dataset	KLd-mixture	KLd-rf	KLd-uniform
1	0.008	0.328	1.936
2	0.008	0.409	2.005
3	0.007	0.266	1.907
4	0.011	0.361	1.963
5	0.005	0.340	1.942
6	0.006	0.390	1.980
7	0.007	0.428	2.011
8	0.006	0.346	1.958
9	0.011	0.314	1.930
10	0.004	0.363	1.951

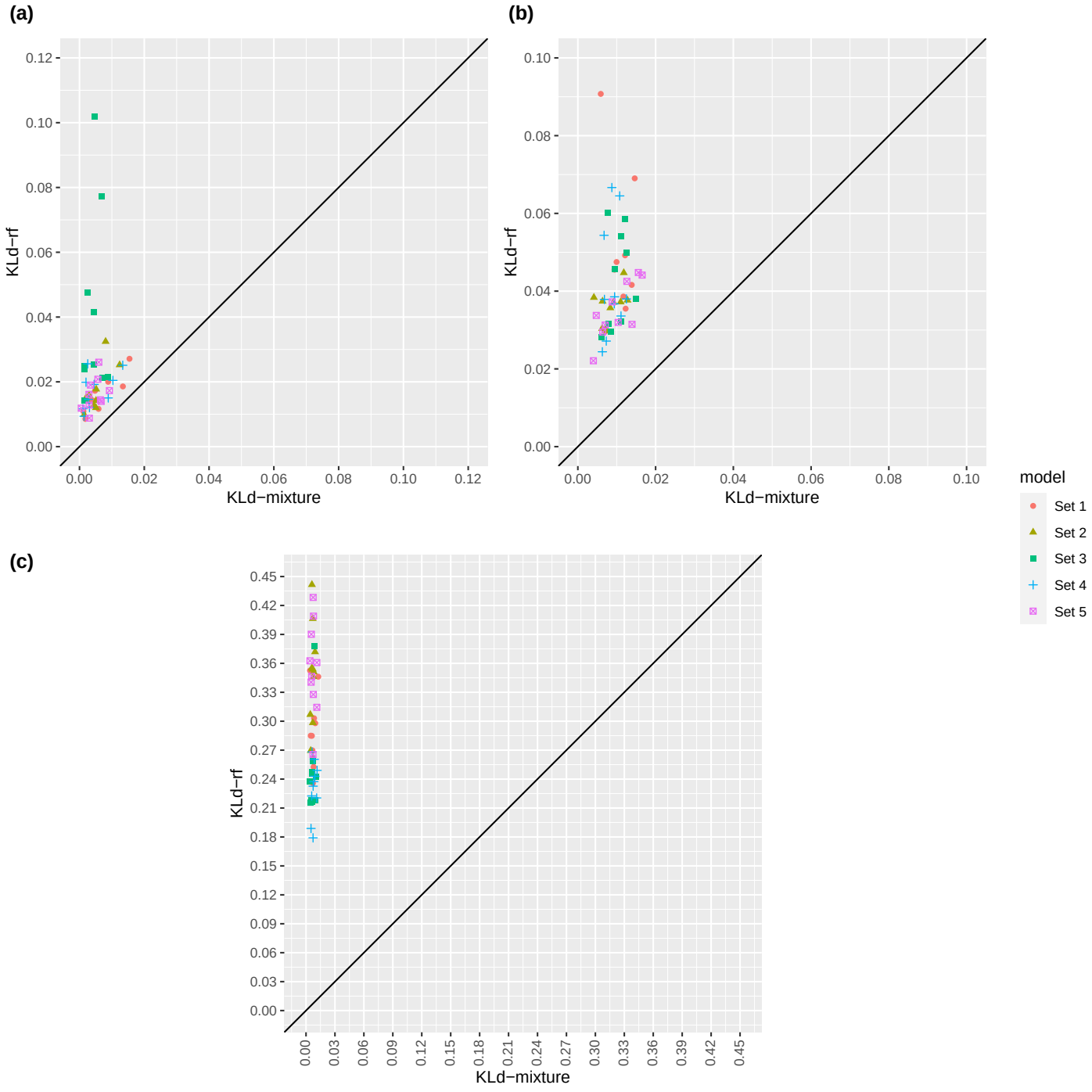


Figure 1: Comparisons of the KL divergences (KLd-mixture) computed between the "true" model probabilities and the probabilities estimated from the proposed mixture model with the KL divergences (KLd-rf) computed between the "true" model probabilities and the relative frequency distribution for each simulation setting (a) $N = 500$, $p = 4$, and $K_{true} = 2$, (b) $N = 1000$, $p = 6$, and $K_{true} = 7$, and (c) $N = 2000$, $p = 8$, and $K_{true} = 10$

B: Teaching Style Study (Section 5.1)

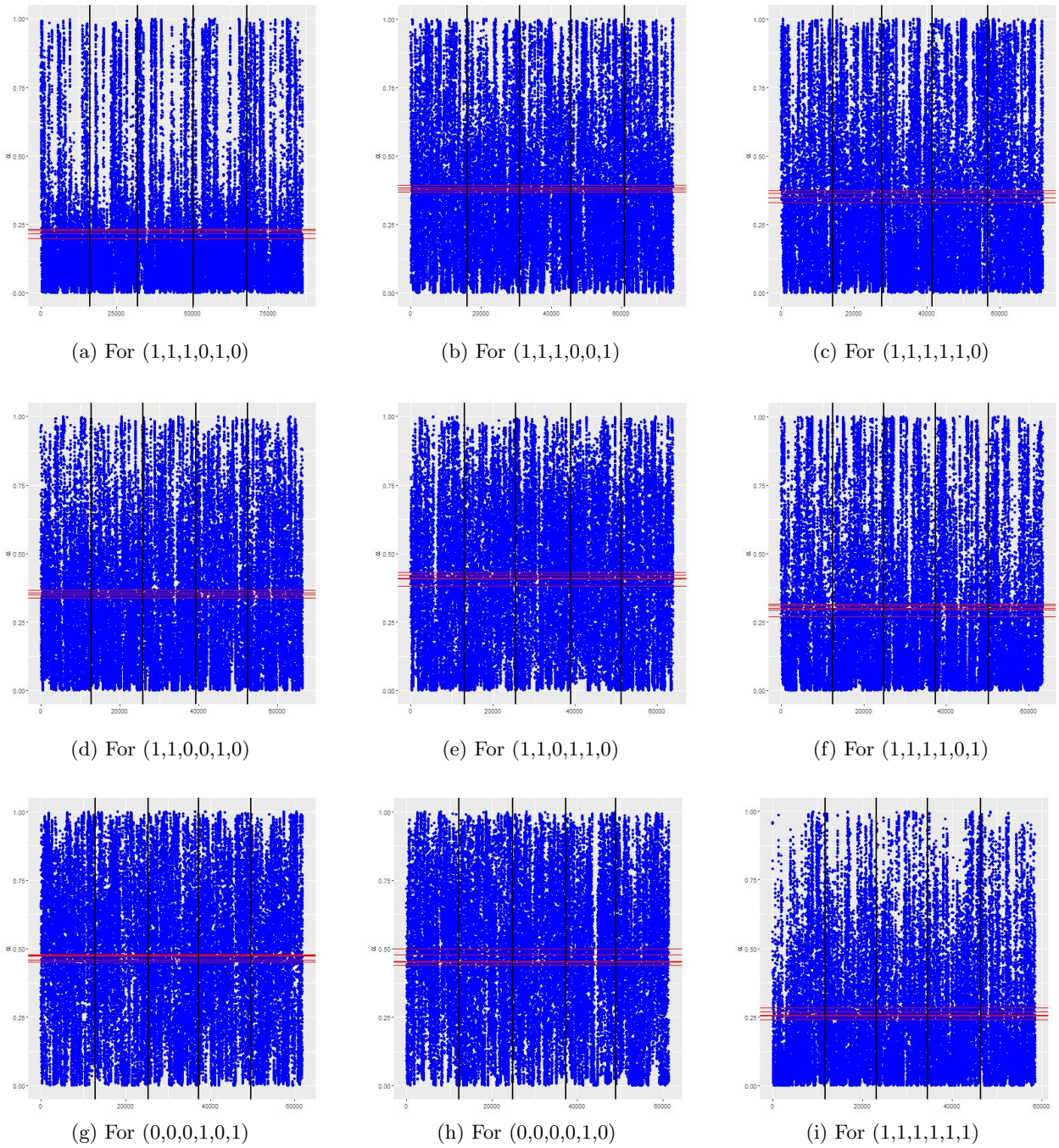
The tables and plots supporting the Teaching Style Study dataset analysis are reported below.

Table 1: Comparison of the relative frequencies of the ten most frequent vectors observed in the Teaching Style Study dataset with the respective probabilities estimated after $T = 10,000$ complete iterations of the first version of the MCMC sampler. The observations are numbered 1 to 10 with 1 denoting the most frequent vector. The five MCMC runs with different starting values are numbered 1 to 5. The top row denotes the relative frequencies observed in the dataset. The KL divergence between the relative frequencies and the estimated probabilities for all observed outcomes is denoted by D_{KL}.

	Outcome Vectors										
Run	1	2	3	4	5	6	7	8	9	10	D_{KL}
—	0.252	0.068	0.056	0.043	0.038	0.032	0.032	0.030	0.026	0.024	—
1	0.251	0.068	0.054	0.042	0.036	0.032	0.030	0.028	0.023	0.025	0.072
2	0.251	0.068	0.054	0.042	0.036	0.032	0.029	0.027	0.023	0.024	0.071
3	0.251	0.068	0.054	0.041	0.036	0.032	0.029	0.028	0.023	0.024	0.074
4	0.251	0.068	0.054	0.041	0.036	0.032	0.028	0.028	0.023	0.024	0.074
5	0.251	0.068	0.054	0.042	0.036	0.032	0.029	0.028	0.024	0.025	0.072

Table 2: Relative frequencies of the number of "non-empty" components used by the MCMC algorithm for each run with the Teaching Style Study dataset.

Run	Number of non-empty components								
—	22	23	24	25	26	27	28	29	30
1	0.0016	0.0094	0.0361	0.0914	0.1903	0.2548	0.2426	0.1403	0.0332
2	0.0055	0.0154	0.0509	0.1105	0.2000	0.2511	0.2215	0.1129	0.0317
3	0.0069	0.0218	0.0604	0.1318	0.2176	0.2482	0.1970	0.0952	0.0201
4	0.0052	0.0177	0.0505	0.1157	0.1973	0.2492	0.2219	0.1151	0.0266
5	0.0027	0.0110	0.0402	0.1009	0.1862	0.2538	0.2362	0.1363	0.0323



(a) For (1,1,1,0,1,0) (b) For (1,1,1,0,0,1) (c) For (1,1,1,1,1,0)

(d) For (1,1,0,0,1,0) (e) For (1,1,0,1,1,0) (f) For (1,1,1,1,0,1)

(g) For (0,0,0,1,0,1) (h) For (0,0,0,0,1,0) (i) For (1,1,1,1,1,1)

Figure 2: Plots for the variability parameter iterates α (all five MCMC runs combined) for the nine most frequent central vectors identified by the mixture model in the Teaching Style Study dataset. For each vector, the iterates are ordered by runs 1 to 5 separated by vertical black lines. The horizontal red lines indicate the mean values for the five runs.

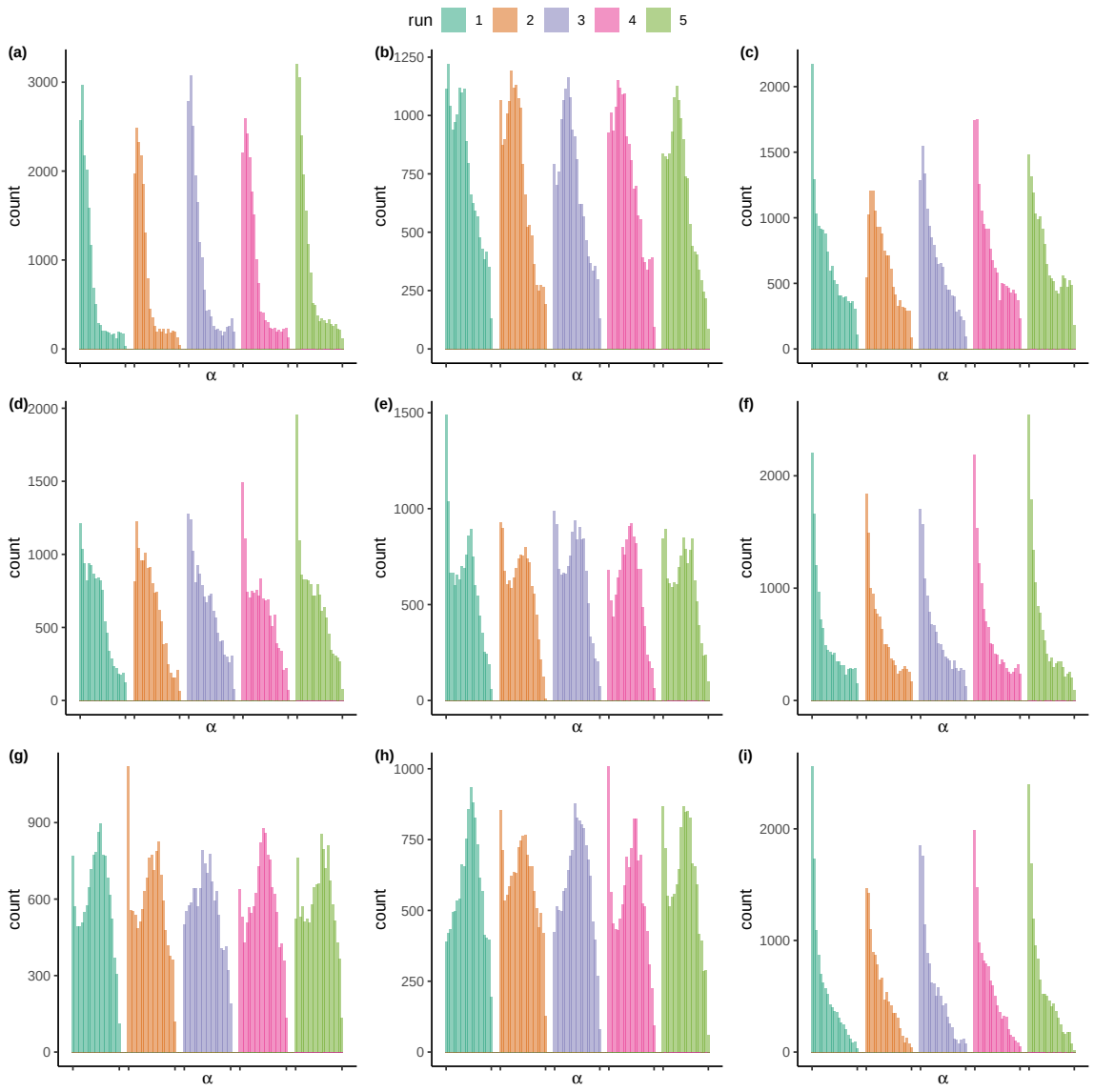


Figure 3: Histograms of the variability parameters α corresponding to the nine most frequent central vectors identified in the Teaching Style Study dataset. The histograms correspond to the vectors (a) $(1,1,1,0,1,0)$, (b) $(1,1,1,0,0,1)$, (c) $(1,1,1,1,1,0)$, (d) $(1,1,0,0,1,0)$, (e) $(1,1,0,1,1,0)$, (f) $(1,1,1,1,0,1)$, (g) $(0,0,0,1,0,1)$, (h) $(0,0,0,0,1,0)$, and (i) $(1,1,1,1,1,1)$. The five runs have been color-coded.

C: Newspaper Reading Survey (Section 5.2)

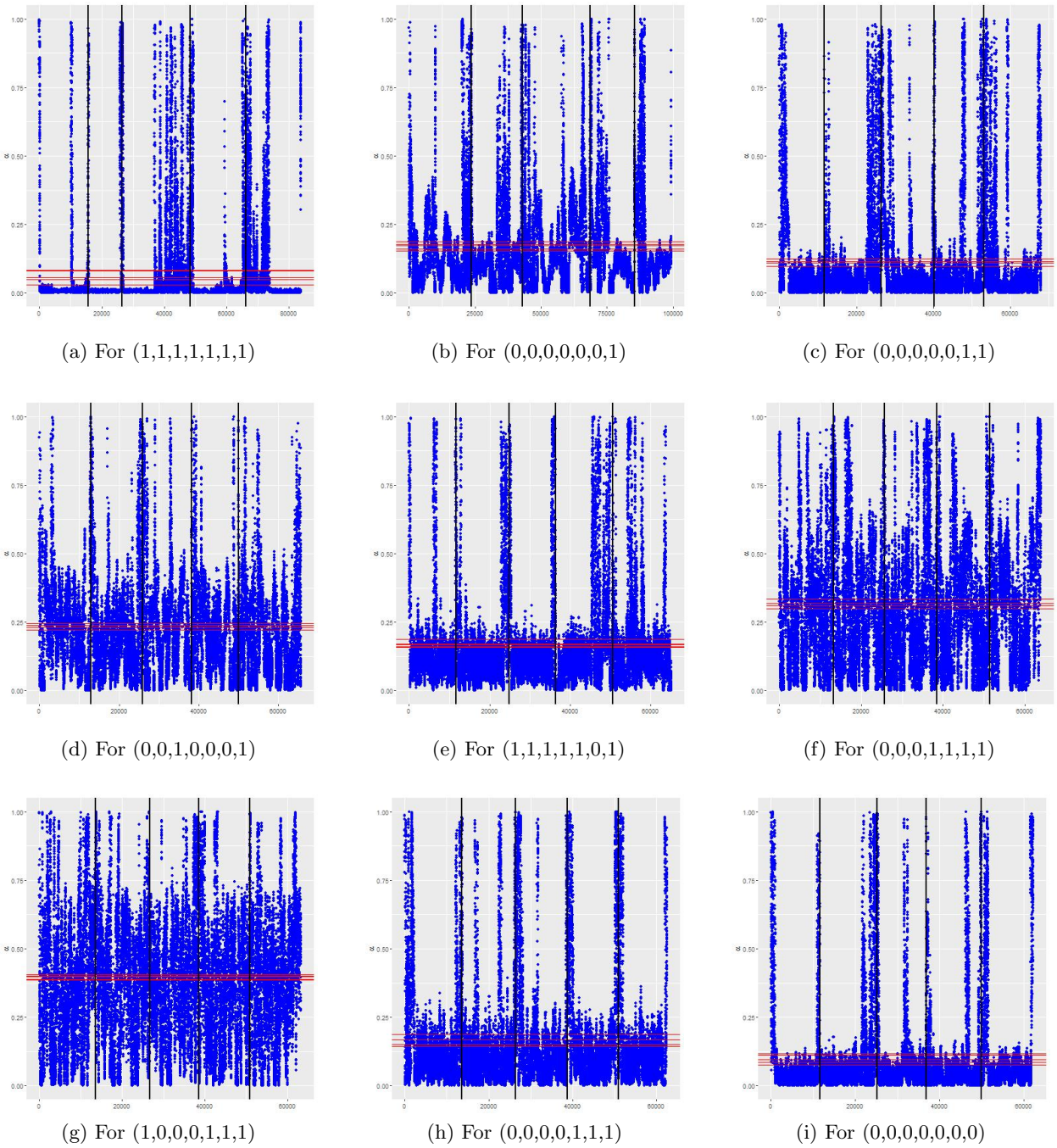
The tables and plots supporting the Newspaper Reading Survey dataset analysis are reported below.

Table 3: Comparison of the relative frequencies of the ten most frequent vectors observed in the Newspaper Reading Survey dataset with the respective probabilities estimated after $T = 10,000$ complete iterations of the second version of the MCMC sampler. The observations are numbered 1 to 10 with 1 denoting the most frequent vector. The five MCMC runs with different starting values are numbered 1 to 5. The top row denotes the relative frequencies observed in the dataset. The KL divergence between the relative frequencies and the estimated probabilities for all observed outcomes is denoted by D_{KL}.

	Outcome Vectors										
Run	1	2	3	4	5	6	7	8	9	10	D_{KL}
—	0.427	0.198	0.048	0.046	0.027	0.027	0.027	0.019	0.014	0.014	—
1	0.426	0.197	0.048	0.046	0.027	0.027	0.027	0.019	0.014	0.014	0.005
2	0.426	0.198	0.048	0.046	0.027	0.027	0.027	0.019	0.014	0.014	0.005
3	0.426	0.198	0.048	0.046	0.027	0.027	0.027	0.019	0.014	0.014	0.005
4	0.426	0.198	0.048	0.046	0.027	0.027	0.027	0.019	0.014	0.014	0.005
5	0.426	0.197	0.048	0.046	0.027	0.027	0.027	0.019	0.014	0.014	0.005

Table 4: Relative frequencies of the number of "non-empty" components used by the MCMC algorithm for each run with the Newspaper Reading Survey dataset.

Run	Number of non-empty components								
—	41	42	43	44	45	46	47	48	49
1	0.0114	0.0303	0.0670	0.1264	0.1845	0.2117	0.1818	0.1155	0.0460
2	0.0122	0.0343	0.0763	0.1367	0.1988	0.2117	0.1793	0.0978	0.0366
3	0.0114	0.0340	0.0738	0.1362	0.2072	0.2215	0.1708	0.0950	0.0294
4	—	0.0311	0.0735	0.1385	0.1882	0.2050	0.1801	0.1033	0.0460
5	0.0123	0.0316	0.0749	0.1353	0.1987	0.2189	0.1751	0.1013	0.0375



(a) For (1,1,1,1,1,1,1)

(b) For (0,0,0,0,0,0,1)

(c) For (0,0,0,0,0,1,1)

(d) For (0,0,1,0,0,0,1)

(e) For (1,1,1,1,1,0,1)

(f) For (0,0,0,1,1,1,1)

(g) For (1,0,0,0,1,1,1)

(h) For (0,0,0,0,1,1,1)

(i) For (0,0,0,0,0,0,0)

Figure 4: Plots for the variability parameter iterates α (all five MCMC runs combined) for the nine most frequent central vectors identified by the mixture model in the Newspaper Reading Survey dataset. For each vector, the iterates are ordered by runs 1 to 5 separated by vertical black lines. The horizontal red lines indicate the mean values for the five runs.

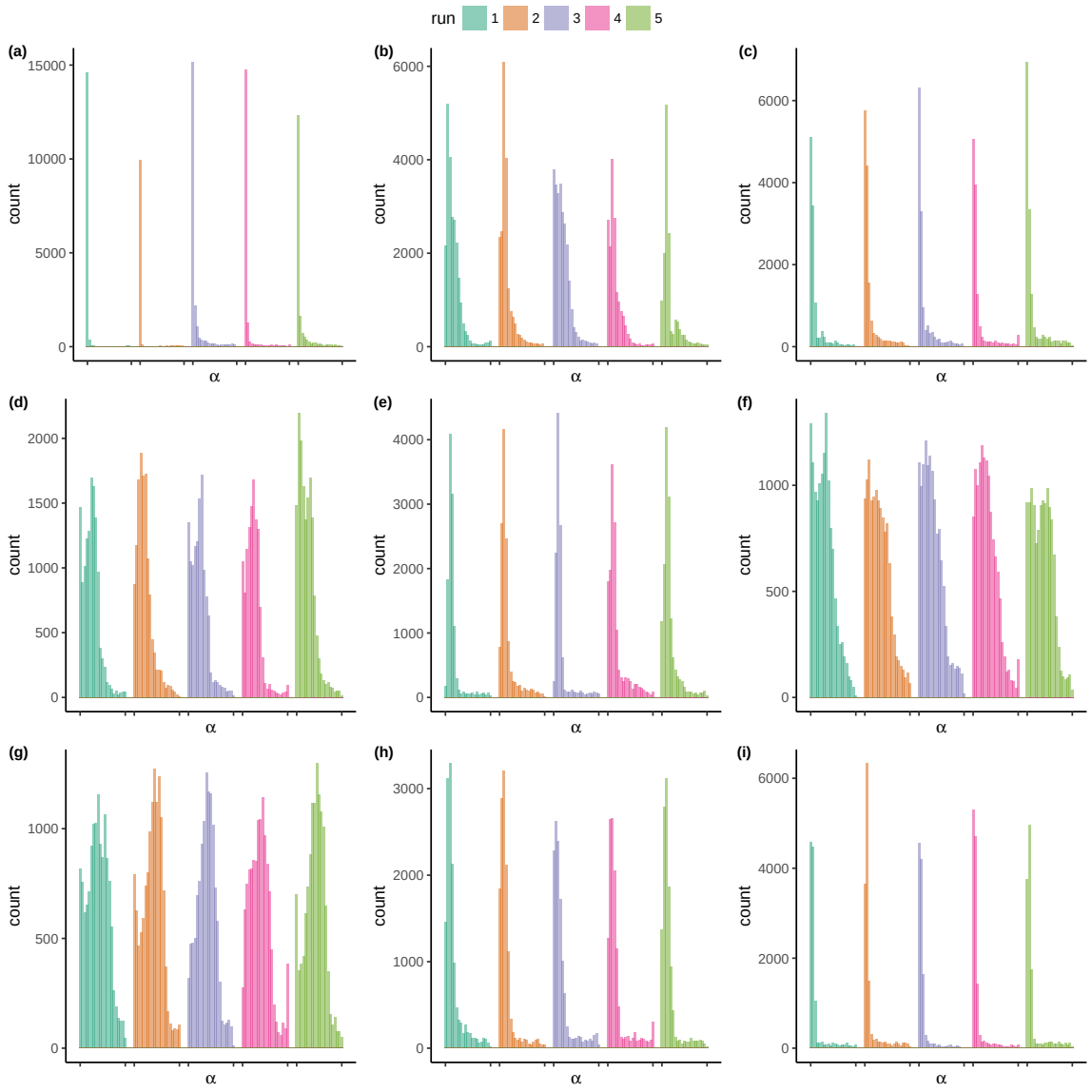


Figure 5: Histograms of the variability parameters α corresponding to the nine most frequent central vectors identified in the Newspaper Reading Survey dataset. The histograms correspond to the vectors (a) $(1,1,1,1,1,1,1)$, (b) $(0,0,0,0,0,0,1)$, (c) $(0,0,0,0,0,1,1)$, (d) $(0,0,1,0,0,0,1)$, (e) $(1,1,1,1,1,0,1)$, (f) $(0,0,0,1,1,1,1)$, (g) $(1,0,0,0,1,1,1)$, (h) $(0,0,0,0,1,1,1)$, and (i) $(0,0,0,0,0,0,0)$. The five runs have been color-coded.

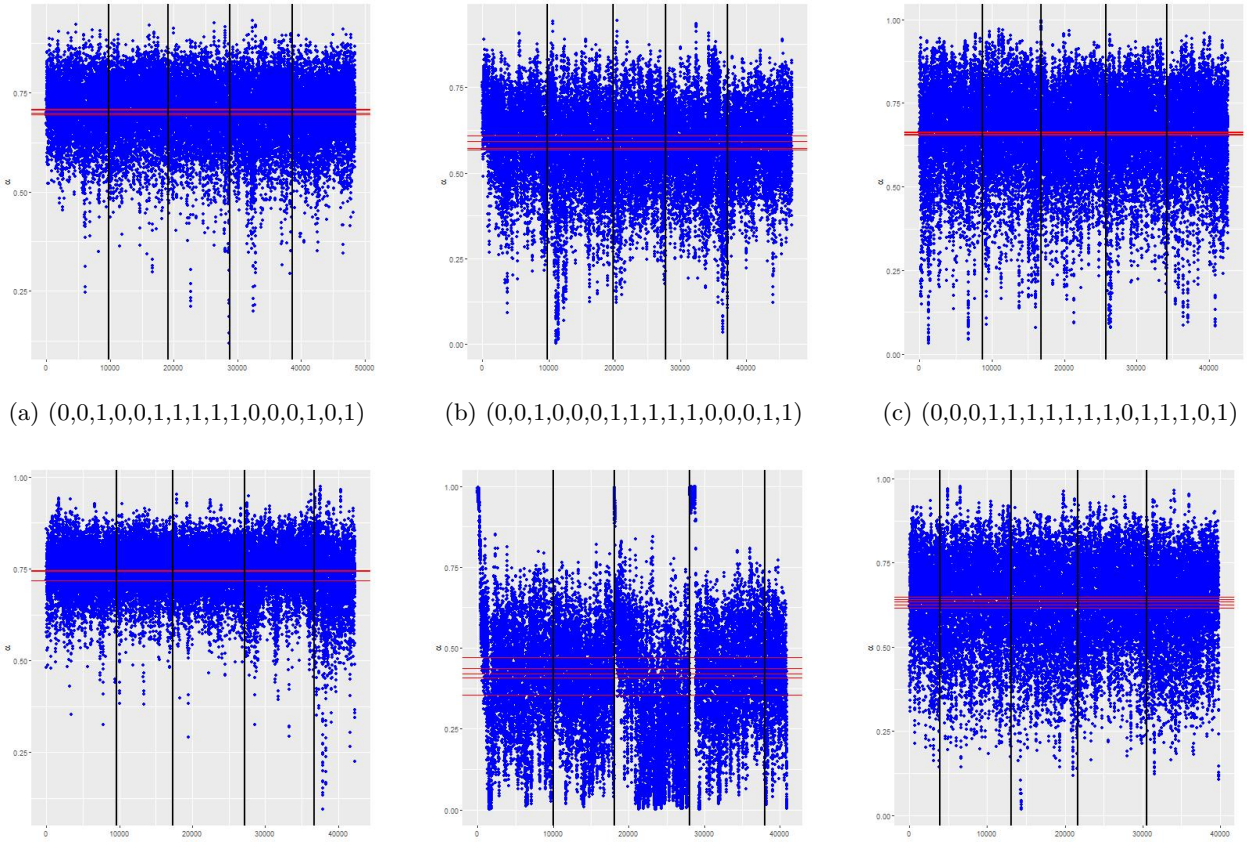
D: HouseVotes84 Dataset (Section 6.3)

The tables and plots supporting the HouseVotes84 dataset analysis are reported below.

Table 5: The KL divergence measures computed, at each MCMC run numbered 1 to 5, between the relative frequencies and the estimated probabilities for the complete data vectors in the full HouseVotes84 dataset and the two datasets defined by 'Class' variable.

Run	Full	Run	Class='democrat'	Run	Class='republican'
1	0.867	1	0.821	1	0.726
2	0.832	2	0.768	2	0.751
3	0.831	3	0.710	3	0.635
4	0.843	4	0.809	4	0.687
5	0.883	5	0.700	5	0.769

Full dataset:



(a) (0,0,1,0,0,1,1,1,1,1,0,0,0,1,0,1)

(b) (0,0,1,0,0,0,1,1,1,1,1,0,0,0,1,1)

(c) (0,0,0,1,1,1,1,1,1,1,1,0,1,1,1,0,1)

(d) (1,1,1,0,1,1,0,0,0,0,1,0,1,1,0,1)

(e) (0,0,0,1,1,1,0,0,0,0,0,1,1,1,0,0)

(f) (0,0,1,0,0,1,1,1,1,0,1,0,0,1,1,1)

Figure 6: Plots for the variability parameter iterates α (all five MCMC runs combined) corresponding to the six most frequent central vectors identified by the mixture model in the HouseVotes84 dataset. For each vector, the values are ordered by runs 1 to 5 separated by vertical black lines. The horizontal red lines indicate the mean values for the five runs.

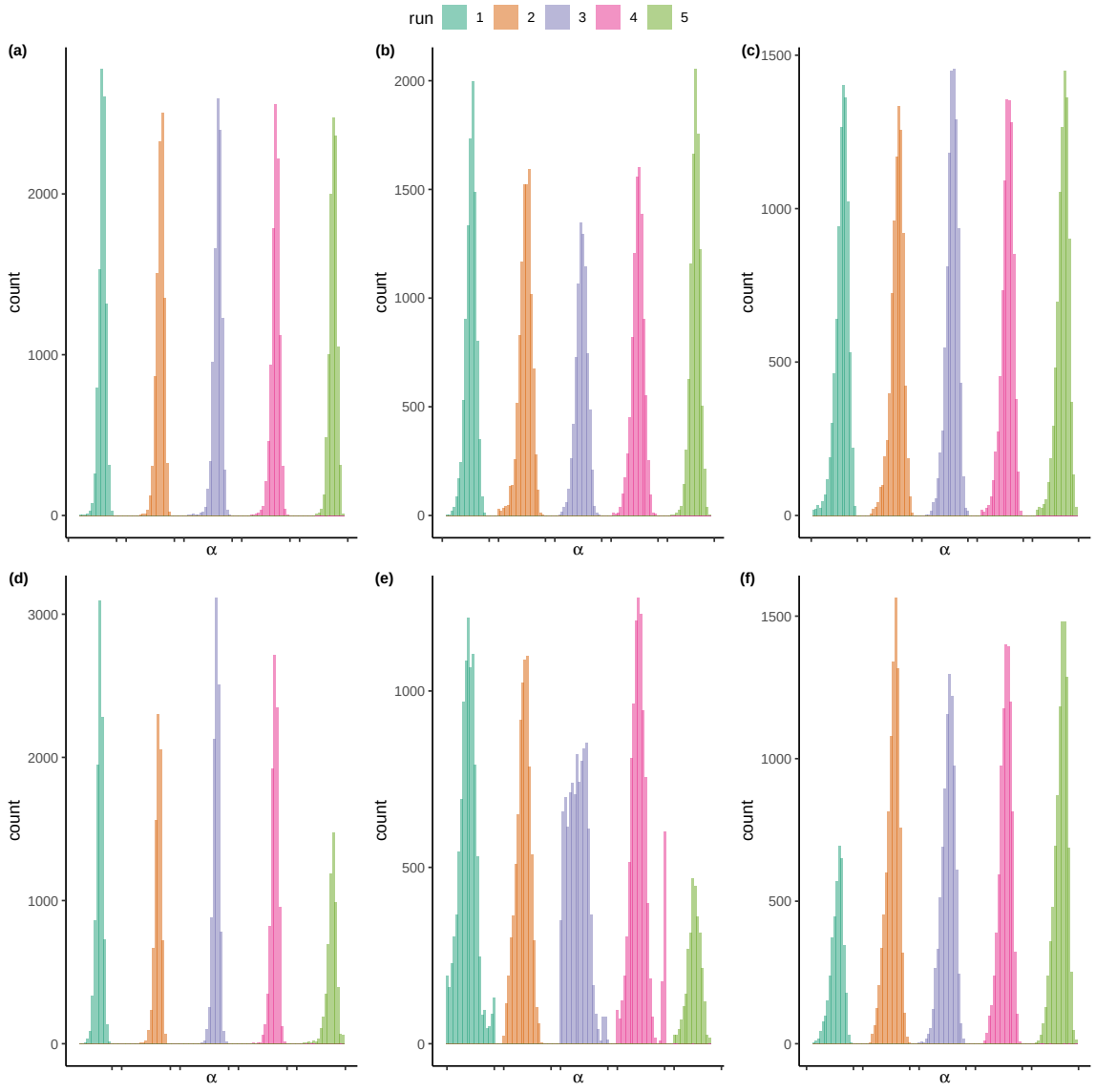
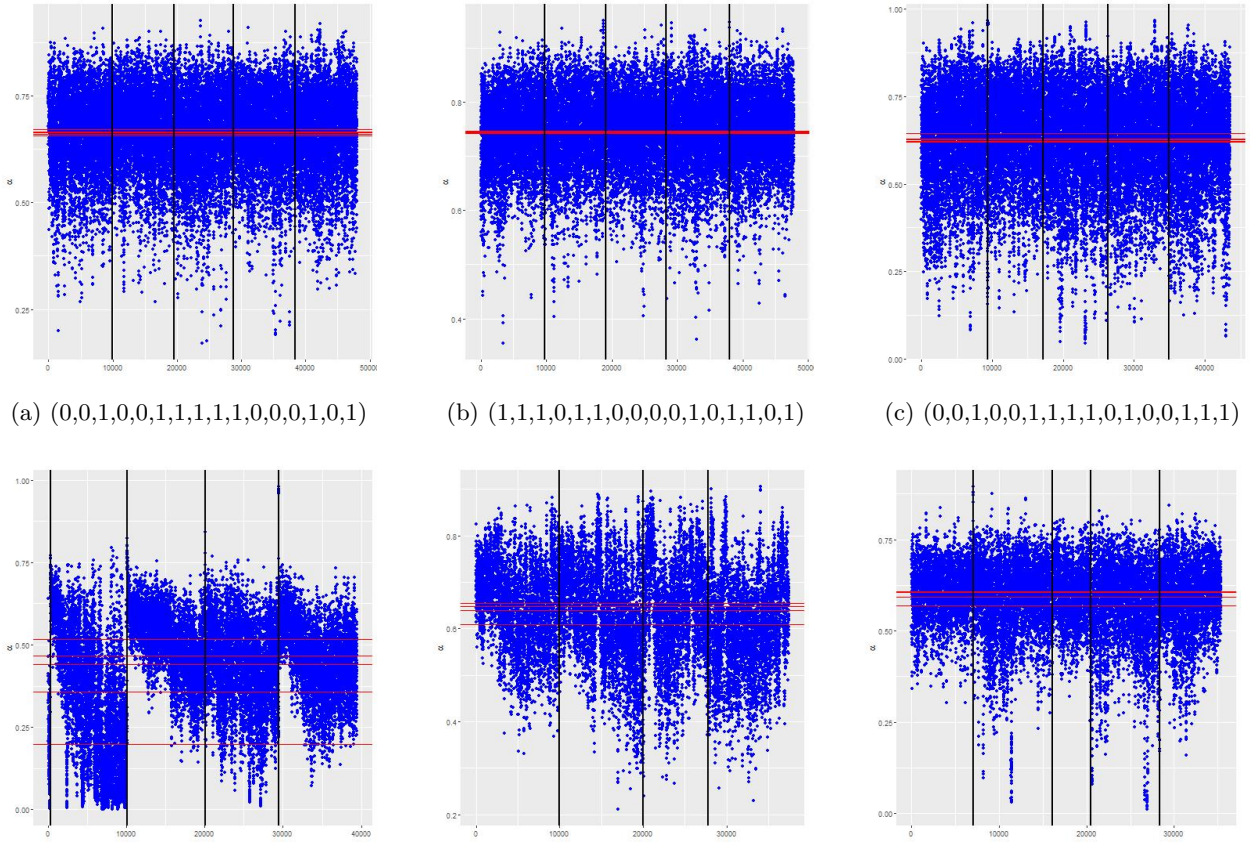


Figure 7: Histograms of the variability parameters α corresponding to the six most frequent central vectors identified in the HouseVotes84 dataset. The histograms correspond to the vectors (a) $(0,0,1,0,0,1,1,1,1,1,0,0,0,1,0,1)$, (b) $(0,0,1,0,0,0,1,1,1,1,1,0,0,0,1,1)$, (c) $(0,0,0,1,1,1,1,1,1,1,0,1,1,1,0,1)$, (d) $(1,1,1,0,1,1,0,0,0,0,1,0,1,1,0,1)$, (e) $(0,0,0,1,1,1,0,0,0,0,0,1,1,1,0,0)$, and (f) $(0,0,1,0,0,1,1,1,1,0,1,0,0,1,1,1)$. The five runs have been color-coded.

Class = 'democrat':



(a) (0,0,1,0,0,1,1,1,1,1,0,0,0,1,0,1)

(b) (1,1,1,0,1,1,0,0,0,0,1,0,1,1,0,1)

(c) (0,0,1,0,0,1,1,1,1,1,0,1,0,0,1,1,1)

(d) (1,1,1,0,0,0,1,1,1,1,0,0,0,0,0,1,1)

(e) (1,0,1,0,0,0,1,1,1,1,1,0,0,0,0,1,1)

(f) (1,1,1,0,0,0,1,1,1,1,0,1,0,0,0,1,1)

Figure 8: Plots for the variability parameter iterates α (all five MCMC runs combined) corresponding to the six most frequent central vectors identified by the mixture model in the HouseVotes84 dataset with Class = 'democrat'. For each vector, the values are ordered by runs 1 to 5 separated by vertical black lines. The horizontal red lines indicate the mean values for the five runs.

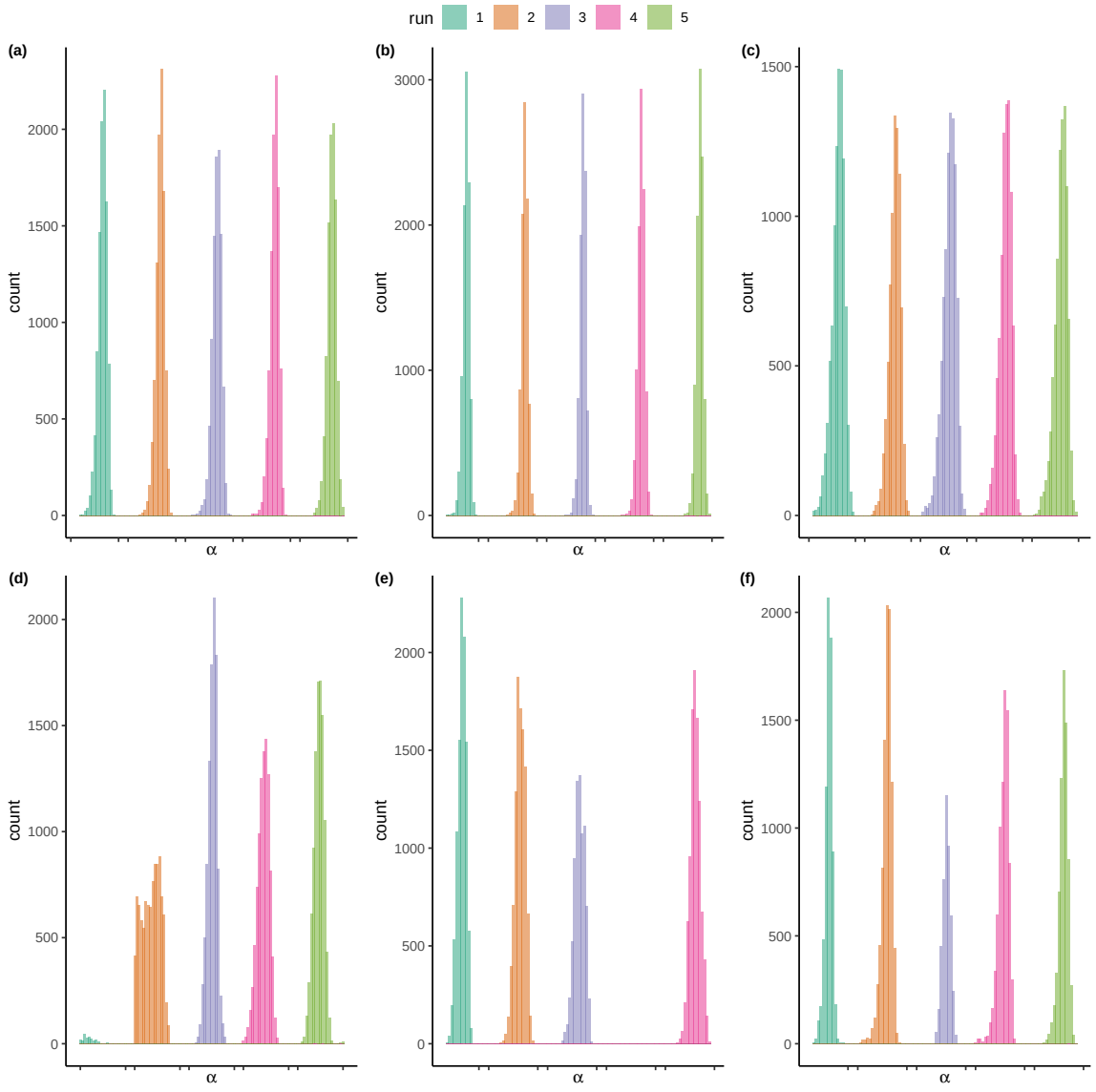
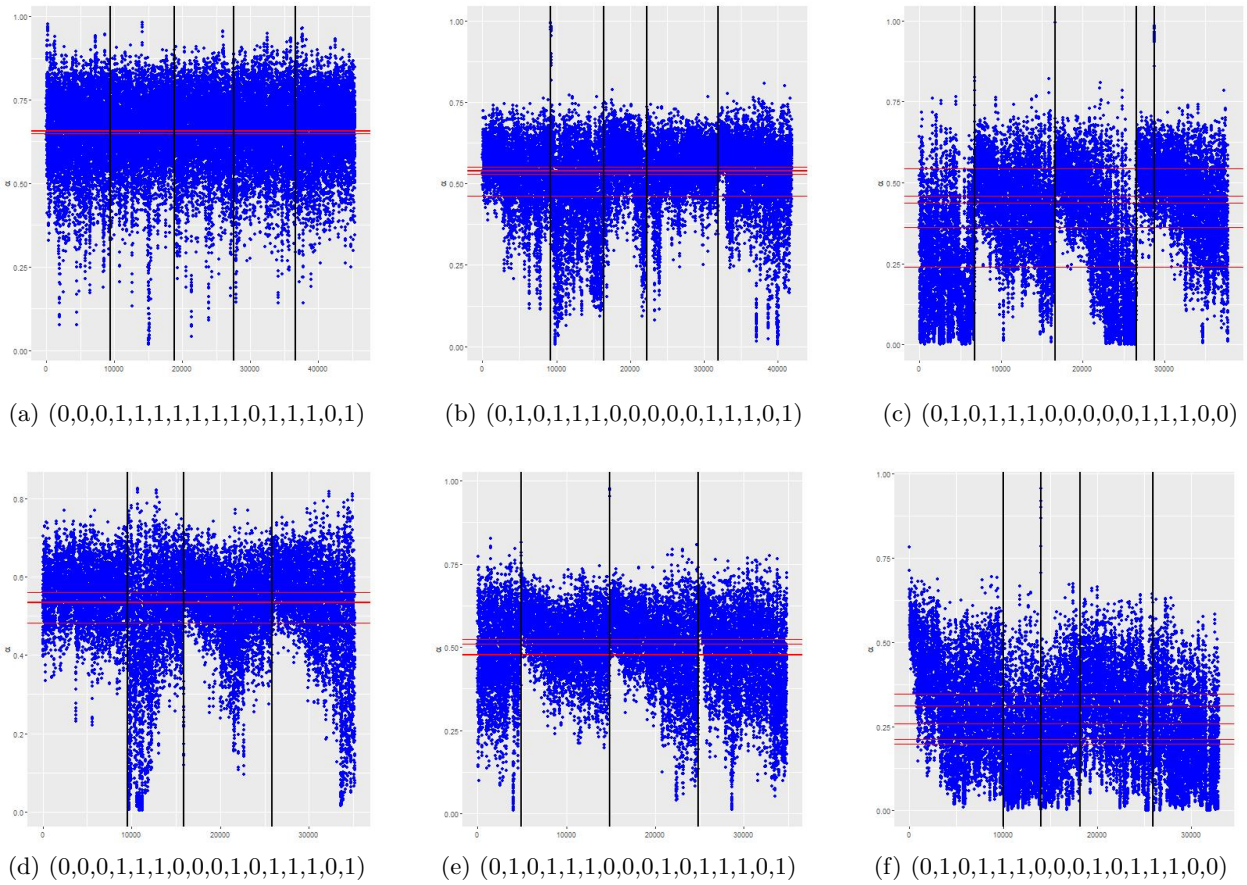


Figure 9: Histograms of the variability parameters α corresponding to the six most frequent central vectors identified in the HouseVotes84 dataset with Class = 'democrat'. The histograms correspond to the vectors (a) $(0,0,1,0,0,1,1,1,1,1,0,0,0,1,0,1)$, (b) $(1,1,1,0,1,1,0,0,0,0,1,0,1,1,0,1)$, (c) $(0,0,1,0,0,1,1,1,1,0,1,0,0,1,1,1)$, (d) $(1,1,1,0,0,0,1,1,1,0,0,0,0,0,1,1)$, (e) $(1,0,1,0,0,0,1,1,1,1,0,0,0,0,1,1)$, and (f) $(1,1,1,0,0,0,1,1,1,0,1,0,0,0,1,1)$. The five runs have been color-coded.

Class = 'republican':



(a) (0,0,0,1,1,1,1,1,1,1,0,1,1,1,0,1)

(b) (0,1,0,1,1,1,0,0,0,0,0,1,1,1,0,1)

(c) (0,1,0,1,1,1,0,0,0,0,0,1,1,1,0,0)

(d) (0,0,0,1,1,1,0,0,0,1,0,1,1,1,0,1)

(e) (0,1,0,1,1,1,0,0,0,1,0,1,1,1,0,1)

(f) (0,1,0,1,1,1,0,0,0,1,0,1,1,1,0,0)

Figure 10: Plots for iterates of the variability parameter α (all five MCMC runs combined) corresponding to the six most frequent central vectors identified by the mixture model in the HouseVotes84 dataset with Class = 'republican'. For each vector, the values are ordered by runs 1 to 5 separated by vertical black lines. The horizontal red lines indicate the mean values for the five runs.

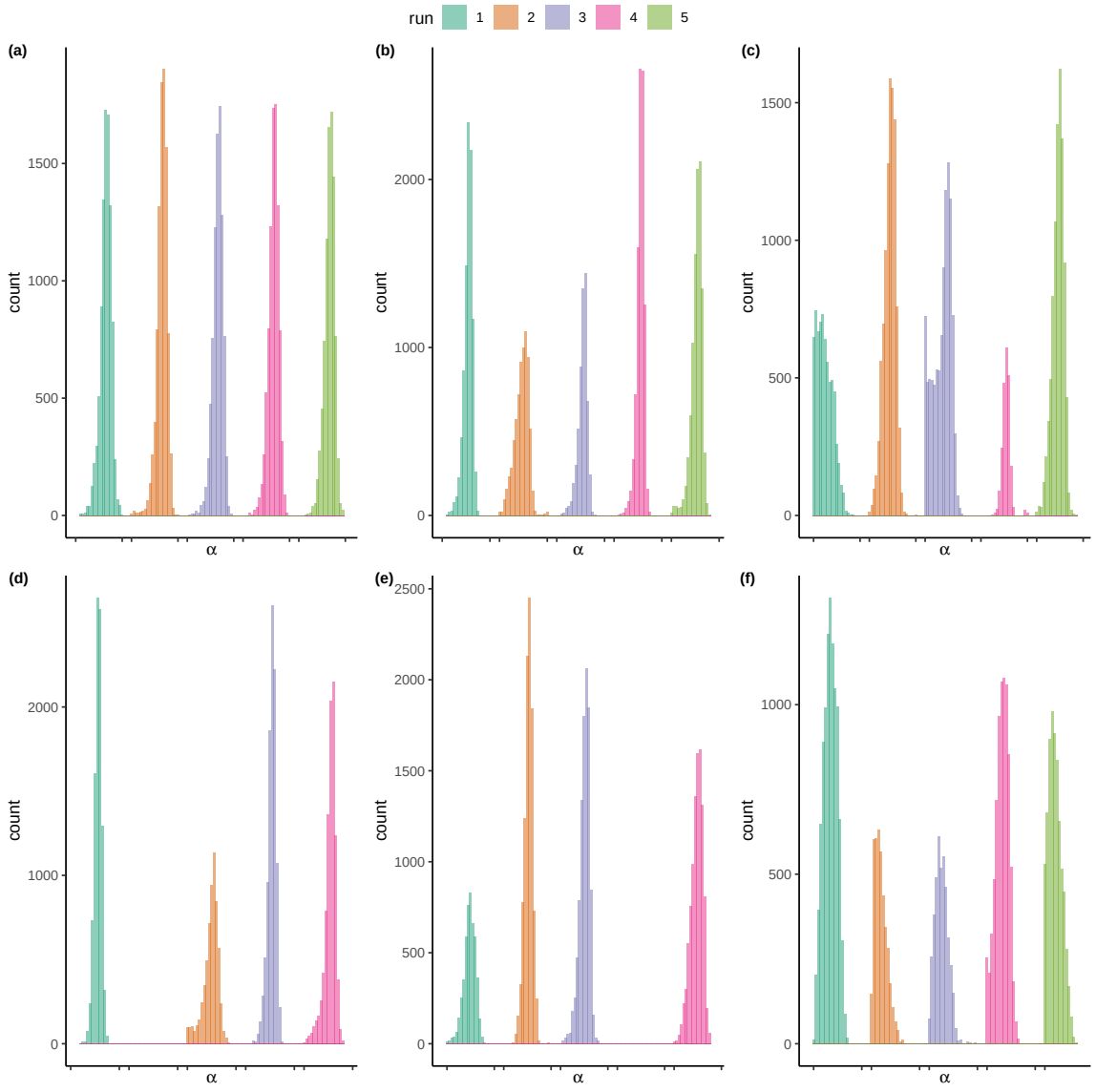


Figure 11: Histograms of the variability parameters α corresponding to the six most frequent central vectors identified in the HouseVotes84 dataset with Class = 'republican'. The histograms correspond to the vectors (a) $(0,0,0,1,1,1,1,1,1,1,0,1,1,1,0,1)$, (b) $(0,1,0,1,1,1,0,0,0,0,0,1,1,1,0,1)$, (c) $(0,1,0,1,1,1,0,0,0,0,0,1,1,1,0,0)$, (d) $(0,0,0,1,1,1,0,0,0,1,0,1,1,1,0,1)$, (e) $(0,1,0,1,1,1,0,0,0,1,0,1,1,1,0,1)$, and (f) $(0,1,0,1,1,1,0,0,0,1,0,1,1,1,0,0)$. The five runs have been color-coded.