-Simulation with n = 200 and K = 8, Kmax = 50

* I tried different heat vectors using the method from Altekar 2004 (table above)
  + All models predicted **K\_map = 7**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Estimated posterior distribution of the number of clusters for each survey wave. | | | | | | | | |
| **Heats vector** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| Temp1 | 0.008 | 0.178 | **0.418** | 0.280 | 0.090 | 0.018 | 0.006 | 0.002 |
| Temp2 | 0.038 | 0.235 | **0.341** | 0.260 | 0.085 | 0.030 | 0.009 | 0.002 |
| Temp3 | 0.091 | 0.222 | **0.332** | 0.219 | 0.104 | 0.025 | 0.005 | 0.002 |
| Temp4 | 0.058 | 0.185 | **0.319** | 0.245 | 0.146 | 0.045 | 0.001 | 0.001 |

|  |  |
| --- | --- |
| Heats vector (4 chains) | Swap acceptance rate |
| 1.0 , 0.9523810, 0.9090909,  0.8695652 | 42.2% |
| 1.0, 0.9090909, 0.8333333 ,0.7692308 | 12.6% |
| 1.0, 0.8695652, 0.7692308, 0.6896552 | 3.5% |
| 1.0, 0.8333333, 0.7142857, 0.6250000 | 0.5% |

**Subset of the ACS 2006-2010 dataset where n= 100, and number of iterations = 100**

|  |  |
| --- | --- |
| Heats vector (4 chains) | Swap acceptance rate |
| 1.0 , 0.9523810, 0.9090909,  0.8695652 | 78% |
| 1.0, 0.9090909, 0.8333333 ,0.7692308 | 40% |
| 1.0, 0.8695652, 0.7692308, 0.6896552 | 20% |
| 1.0, 0.8333333, 0.7142857, 0.6250000 | 14% |

****

I explored the output for the highlighted model. First, I looked at the sampled K values per heated chain, where chain 1 corresponds to the target posterior.

**A graph with colored lines

Description automatically generated**

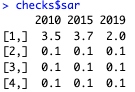
I looked at the raw values of the mixture weights and then the reordered according to ECR algorithm to see if samples were mixing well using this vector of heats. It looks good, even with a small sample and number of iterations.

**A graph with red and blue dots

Description automatically generated**

Notes

* **From the latest tuning of the heats vector, it seems like smaller deltaT work better for these datasets.**



* We are keeping the model with the proportion of Hispanic, NHB and NHW
* We are going to “gray” out clusters that are not very populated.
  + This will be noted in the bar charts and the maps.
* Updated
  + Cluster profiles.
    - The only external data I looked at was the age distribution since now we have included more variables.
    - Table with census variables description and distribution as binary variables