# 1 Results

Our clustering analysis was conducted on 100 data points, the former 50 of which were iid sampled from a $\mathcal{N}(4, 1)$ and the latter from a $\mathcal{N}(7, 1)$, which were saved in the `data.csv` file. We chose the prior parameters for the Normal-NIG model **??** as follows: $\mu_0 = 5, \lambda = 1, \alpha_0 = 2, \beta_0 = 2$. The `Neal8` algorithm with $m = 3$ auxiliary blocks was run for 20000 iterations, and the first 5000 were discarded as burn-in. The obtained best clusterings (again, in the lest square sense) and local density estimates are highly fluctuating over the iterations of the algorithm, as shown by the below plots:
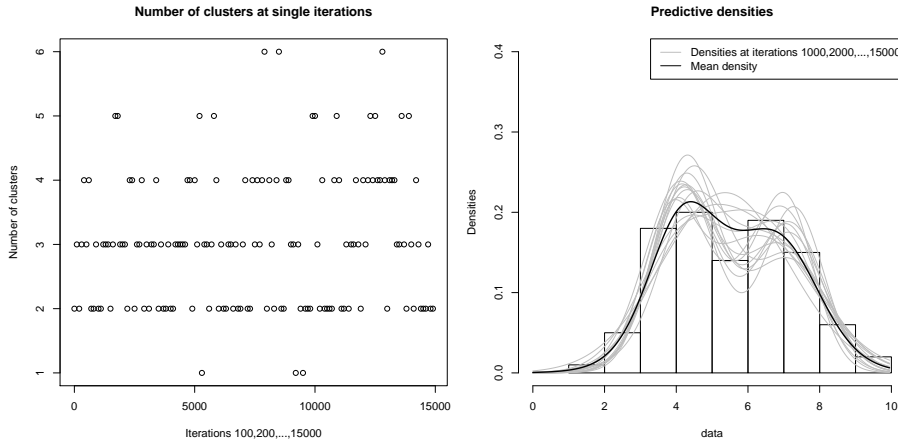

In both plots, a thinning of one iteration every 100 and every 2500, respectively, was performed for better readability of the plot. In the right side plot, the local densities are compared with the barplot of the data as well as the final estimate provided by the mean density. We can see that the number of clusters at all iterations varies significantly between 1 and 6, even in the last thousands of iterations, and the same behavior applies to the local densities. This is further confirmation of the fact that the single iterations themselves do not converge, but instead, as previously discussed, the convergence is in the mean, both for the density estimate and for the average dissimilarity matrix which we use to find the best clustering.

Let us now examine the role of the total mass parameter, M. We run the algorithm with several values for M whilst keeping the other parameters unchanged from the ones indicated at the beginning of the section. For each M, we saved the number of clusters of the best clustering produced by the algorithm, and plotted it against the corresponding M:

Number of clusters as a function of the total mass
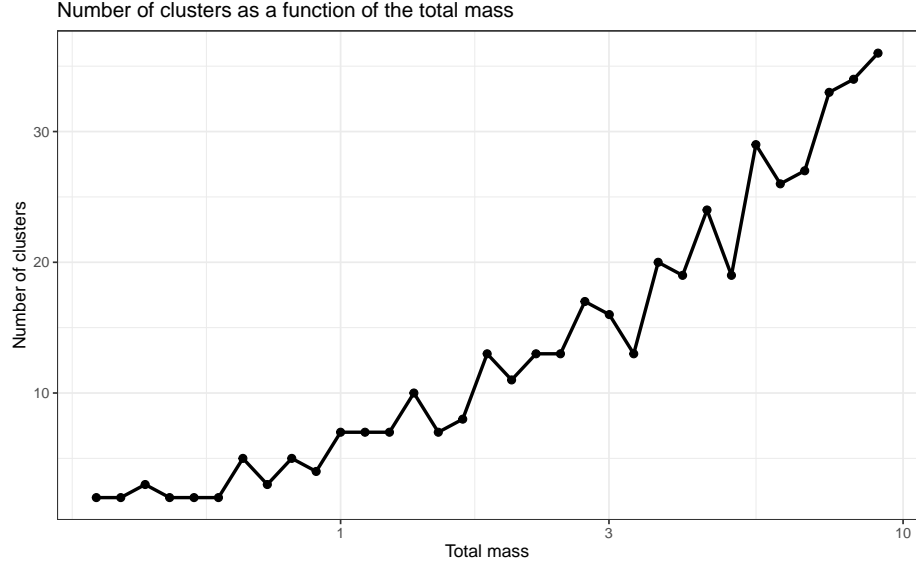
Note that the values for M were chosen so as to be evenly spaced in log-scale, thus the abscissa is in log-scale as well. We can note that the clusters are increasing with the total mass. This is consistent with the fact that the probability of creating a new cluster is proportional to M, as seen in **??**.
Moreover, the density estimates for some of the values of M (again, compared with the barplot of the data) are as follows:

Estimated posterior density
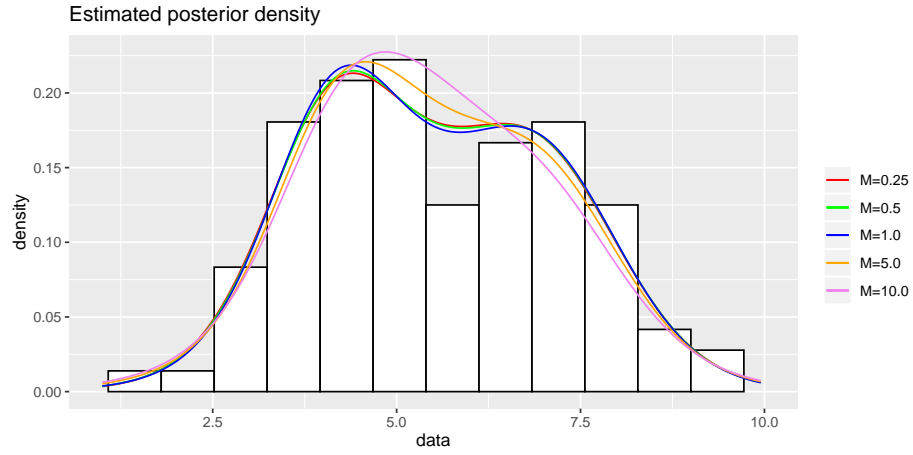
Lower values for the total mass account better for the distribution of the data points, with the modes being near the real expected values of the two normal distributions, 4 and 7. On the other hand, higher values tend to clump together all 100 data points as though they were extracted from a single distribution. This is because TODO

We shall now try and change the number of auxiliary blocks, m, and check how this impacts the the density estimation. For this test, a large total mass

$M = 10$ was chosen; the reason being that a small M would not allow significant differences as m changes. Indeed, the only direct impact of m is in the estimation of the marginal (see **??**), which has a weight of $\frac{M}{M+n}$, which is small if M is also small. The result was the following: this part is super WIP

Estimated posterior density
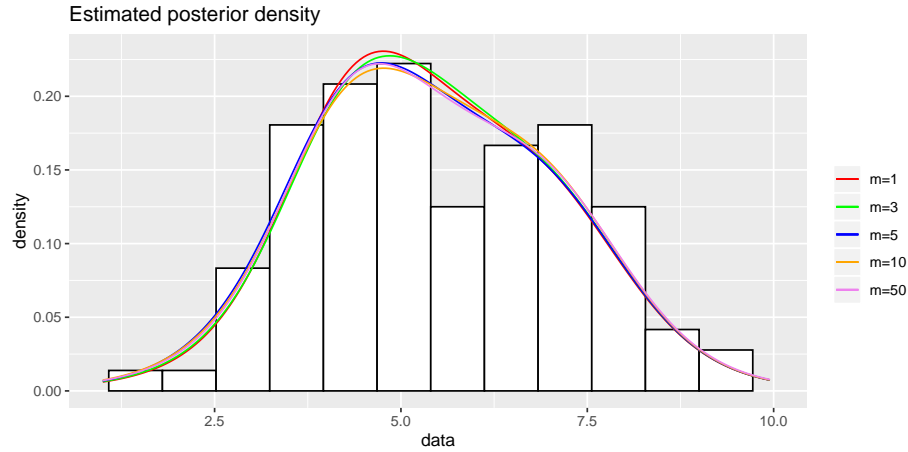
If not explicitly mentioned, everything is Neal8.

3-4) Density estimate as m changes, with M=10 * The estimate for the marginal becomes better (sample mean over a higher number terms) * Doesn't change much with low M b/c the marginal has minuscle weight in the density estimation * We will choose m=3

1) component weights with M=1, m=3 * higher number of clusters than other values * all clusters bar the first two are insignificant (b/c the weights are small and thus...) 2) Density components and full density at best iteration (2490) for M=0.25, m=3

6) Density estimate Neal2 vs Neal8, with m=3, M=0.25 * again, similar estimates at low M b/c the marginal is almost irrelevant