STAT 222 Homework #3 **Mary Silva**
UCSC 18 May 2020

2. Consider the location normal DP mixture model:

$$f(y|G,\phi) = \int k_N(y|\theta,\phi)dG(\theta), \qquad G|\alpha,\mu,\tau^2 \sim DP(\alpha, G_0 = N(\mu,\tau^2))$$

Where $k_N(\cdot|\theta,\phi)$ denotes the density function of the normal distribution with mean θ and variance ϕ. We assume an $IG(a_\phi, b_\phi)$ prior for ϕ, and $Gamma(a_\alpha, b_\alpha)$ prior for α, and take $N(a_\mu, b_\mu)$ and $IG(a_{\tau^2}, b_{\tau^2})$ priors for the mean and variance, μ and τ^2, respectively, for the centering distribution G_0. Therefore, the hierarchical version of this semiparametric DP mixture model is given by

$$\begin{aligned} y_i|\theta_i,\phi &\overset{ind}{\sim} N(y_i|\theta_i,\phi), i = 1,...,n \\ \theta_i|G &\overset{iid}{\sim} G, i = 1,...,n \\ G|\alpha,\mu,\tau^2 &\sim DP(\alpha, G_0 = N(\mu,\tau^2)) \\ \alpha,\mu,\tau^2,\phi &\sim p(\alpha)p(\mu)p(\tau^2)p(\phi) \end{aligned}$$

with the independent priors $p(\alpha), p(\mu), p(\tau^2)$, and $p(\phi)$ for α,μ,τ^2,ϕ given above.

(2.1) **Expressions for the full conditionals.** We derive the expressions for the full conditionals of the Pòlya urn based Gibbs sampler, which can be used for posterior simulation from $p(\boldsymbol{\theta},\alpha,\phi,\mu,\tau^2|data)$, where $data = \{y_i : i = 1,...,n\}$.

A key property for the implementation of the Gibbs sampler is the discreteness of G, which includes a partitioning of the θ_i. From the lecture notes we use the following notation

- n^*: the number of distinct elements(clusters) in the vector $(\theta_1,...,\theta_n)$.
- $\theta_j^*, j = 1,...,n^*$: the distinct θ_i
- $\boldsymbol{w} = (w_1,...,w_n)$ is the vector of configuration indicators, defined by $w_i = j$ if and only if $\theta_i = \theta_j^*, i = 1,...,n$
- n_j is the size of the j^{th} cluster, i.e. $n_j = |\{i : w_i = j\}|, j = 1,...,n^*$.

The vectors $(n^*, \boldsymbol{w}, \theta_1^*,...,\theta_{n^*}^*)$ and $(\theta_1,...,\theta_n)$ are equivalent.

For each $i = 1,...,n$, $p(\theta_i|\{\theta_{i'} : i' \neq i\},\alpha,\mu,\tau^2,\phi,\boldsymbol{y})$ is simply a mixture of n^{*-} point masses and the posterior for θ_i based on y_i,

$$\frac{\alpha q_0}{\alpha q_0 + \sum_{j=1}^{n*^-} n_j^- q_j} h(\theta_i|\mu,\tau^2,\phi,y_i) + \sum_{j=1}^{n*^-} \frac{n_j^- q_j}{\alpha q_0 + \sum_{j=1}^{n*^-} n_j^- q_j} \delta_{\theta_j^*}(\theta_i)$$

where

- $q_j = k_N(y_i|\theta_j^*,\phi)$
- $q_0 = \int k_N(y_i|\theta_i,\phi)g_0(\theta|\mu,\tau^2)d\theta$
- $h(\theta_i|\mu,\tau^2,\phi,y_i) \propto k_N(y_i|\theta_i,\phi)g_0(\theta_i|\mu,\tau^2)$
- g_0 is the density of $G_0 = N(\cdot|\mu,\tau^2)$
- The superscript "-" denotes all relevant quantities when θ_i is removed from the vector $\boldsymbol{\theta}$, e.g. $n*^-$ is the number of clusters in $\{\theta_{i'} : i' \neq i\}$

Integrating out θ, we get the following

$$
\begin{aligned}
q_0 &= \int k_N(y_i|\theta_i,\phi) g_0(\theta|\mu,\tau^2) d\theta \\
&= \int (2\pi\phi)^{-1/2} \exp\left\{-\frac{1}{2\phi}(y_i-\theta)^2\right\} \left(2\pi\tau^2\right)^{-1/2} \exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta \\
&= \left(4\pi^2\phi\tau^2\right)^{-1/2} \int \exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta \\
&= \left(4\pi^2\phi\tau^2\right)^{-1/2} \int \exp\left\{-\frac{1}{2\phi\tau^2}\left[\tau^2 y_i - 2y_i\tau^2\theta + \tau^2\theta^2 + \phi\mu^2 - 2\mu\phi\theta + \phi\theta^2\right]\right\} d\theta \\
&= \left(4\pi^2\phi\tau^2\right)^{-1/2} \int \exp\left\{\frac{1}{2\phi\tau^2}\left[\theta^2(\phi+\tau^2) - 2\theta(\mu\phi + y_i\tau^2)\right] - \frac{\phi\mu^2+\tau^2 y_i^2}{2\phi\tau^2}\right\} d\theta \\
&= \left(4\pi^2\phi\tau^2\right)^{-1/2} \exp\left\{-\frac{\phi\mu^2+\tau^2 y_i^2}{2\phi\tau^2}\right\} \\
&\qquad \times \int \exp\left\{-\frac{\phi+\tau^2}{2\phi\tau^2}\left[\theta^2 - 2\theta\left(\frac{\mu\phi+y_i\tau^2}{\phi+\tau^2}\right) + \left(\frac{\mu\phi+y_i\tau^2}{\phi+\tau^2}\right)^2 - \left(\frac{\mu\phi+y_i\tau^2}{\phi+\tau^2}\right)^2\right]\right\} d\theta \\
&= \left(2\pi\phi\tau^2\right)^{-1/2} \left(\frac{\phi\tau^2}{\phi+\tau^2}\right)^{1/2} \exp\left\{-\frac{\phi\mu^2+\tau^2 y_i}{2\phi\tau^2} + \frac{\left(\mu\phi+y_i\tau^2\right)^2}{2\phi\tau^2(\phi+\tau^2)}\right\} \\
&= \left(2\pi(\phi+\tau^2)\right)^{-1/2} \exp\left\{\frac{(\phi\mu^2+\tau^2 y_i^2)(\phi+\tau^2) + (\mu\phi+y_i\tau^2)^2}{2\phi\tau^2(\phi+\tau^2)}\right\} \\
&= \left(2\pi(\phi+\tau^2)\right)^{-1/2} \exp\left\{\frac{-\phi\mu\tau^2 - \phi y_i\tau^2 + 2\mu\phi\tau^2}{2\phi\tau^2(\phi+\tau^2)}\right\} \\
&= \left(2\pi(\phi+\tau^2)\right)^{-1/2} \exp\left\{-\frac{y_i^2 - 2y_i\mu + \mu^2}{2(\phi+\tau^2)}\right\} \\
&= N(y_i|\mu, \phi+\tau^2)
\end{aligned}
$$

After dropping all the non θ_i terms, we are left with the part that was inside the integral, which is just $h(\theta_i|\cdot)$, which has a normal distribution with mean $\frac{(\mu\phi+y_i\tau^2)}{\phi+\tau^2}$, and variance $\frac{\phi\tau^2}{\phi+\tau^2}$. This completes the marginal posterior for θ_i.
We derive the full conditionals for μ, τ^2, and ϕ:

$$
\begin{aligned}
p(\mu|\cdot) &\propto \exp\left\{-\frac{1}{2b_\mu}(\mu-a_\mu)^2\right\} \exp\left\{-\frac{1}{2\tau^2}\sum_{j=1}^{n^*}(\theta_j^*-\mu)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2b_\mu}(\mu^2 - 2a_\mu\mu + a_\mu^2) - \frac{1}{2\tau^2}\left(\sum_{j=1}^{n^*}\theta_j^{*2} - 2\mu\sum_{j=1}^{n^*}\theta_j^* + n^*\mu^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2b_\mu}(\mu^2 - 2a_\mu\mu) - \frac{1}{2\tau^2}\left(-2\mu\sum_{j=1}^{n^*}\theta_j^* + n^*\mu^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2b_\mu\tau^2}\left(\mu^2\tau^2 - 2a_\mu\mu\tau^2 - 2\mu\sum_{j=1}^{n^*}\theta_j^* b_\mu + \mu^2 n^* b_\mu\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2b_\mu\tau^2}\left(\mu^2(\tau^2+b_\mu n^*)\right) - 2\mu\left(a_\mu\tau^2 + b_\mu\sum_{j=1}^{n^*}\theta_j^*\right)\right\} \\
&\propto \exp\left\{-\frac{\tau^2+b_\mu n^*}{2b_\mu\tau^2}\left[\mu^2 - 2\mu\left(\frac{a_\mu\tau^2 + b_\mu\sum_{j=1}^{n^*}\theta_j^*}{\tau^2+b_\mu n^*}\right)\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
p(\tau^2|\cdot) &\propto (\tau^2)^{a_{\tau^2}-1} \exp\left\{-\frac{b_{\tau^2}}{\tau^2}\right\} (\tau^2)^{n^*/2} \exp\left\{-\frac{1}{2\tau^2}\sum_{j=1}^{n^*}(\theta_j^* - \mu)^2\right\} \\
&= (\tau^2)^{-(a_{\tau^2}+n^*/2)-1} \exp\left\{-\frac{1}{\tau^2}\left(b_{\tau^2} + \frac{1}{2}\sum_{j=1}^{n^*}(\theta_j^* - \mu)^2\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
p(\phi|\cdot) &\propto p(\phi)\prod_{i=1}^{n} k_N(y_i|\theta_i, \phi) \\
&\propto \phi^{-a_\phi-1} \exp\left\{-\frac{b_\phi}{\phi}\right\}\prod_{i=1}^{n}\phi^{-1/2}\exp\left\{-\frac{1}{2\phi}(y_i-\theta_i)^2\right\} \\
&\propto \phi^{-(a_\phi+\frac{n}{2})-1}\exp\left\{-\frac{1}{\phi}\left(b_\phi + \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_i)^2\right)\right\}
\end{aligned}
$$

Thus, after rearranging the terms, the full conditionals for μ, τ^2, and ϕ reduce to

$$
\begin{aligned}
\mu|\cdot &\sim N\left(\left(\frac{1}{b_\mu}+\frac{n^*}{\tau^2}\right)^{-1}\left(\frac{a_\mu}{b_\mu}+\frac{1}{\tau^2}\sum_{j=1}^{n^*}\theta_j^*\right), \left(\frac{1}{b_\mu}+\frac{n^*}{\tau^2}\right)^{-1}\right) \\
\tau^2|\cdot &\sim IG\left(a_{\tau^2}+\frac{n^*}{2}, b_{\tau^2}+\frac{1}{2}\sum_{j=1}^{n^*}(\theta_j^*-\mu)^2\right) \\
\phi|\cdot &\sim IG\left(a_\phi+\frac{n}{2}, b_\phi+\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_i)^2\right)
\end{aligned}
$$

The full conditional for α is not standard form. To estimate, we use the exchangeable product partition function (Antoniak, 1974)

$$
\begin{aligned}
p(n^*|\alpha) &\propto \alpha^{n^*}\frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \\
&= \frac{\alpha^{n^*}(\alpha+n)}{\alpha\Gamma(n)}\int_0^1 \eta^\alpha(1-\eta)^{n-1}d\eta
\end{aligned}
$$

Therefore, we devise a sampler for α by first introducing latent variable η such that

$$\eta|\alpha, n^*, \boldsymbol{y} \sim Beta(\alpha+1, n)$$

Then $\alpha|\eta, n^* \sim \varepsilon Gamma\left(a_\alpha + n^*, b_\alpha - \log\eta\right) + (1-\varepsilon)Gamma\left(a_\alpha+n^*-1, b_\alpha-\log\eta\right)$, where

$$\varepsilon = \frac{a_\alpha+n^*-1}{n(b_\alpha-\log\eta)+a_\alpha+n^*-1}$$

This clever auxiliary variable sampler was first introduced in Escobar and West (1995).

(2.2) **Prior specification.** We examine several choices for the hyperparameters for ϕ, τ^2 and μ. To do this, I chose hyperparameters in such a way that the prior means for each ϕ, τ^2 and μ ranged from small to large and observed how the posterior means for ϕ, τ^2 and μ changed.

For instance, in Figure 1, I kept the priors for τ^2 and ϕ the same while changing the hyperparameters for μ. The posterior mean for μ didn't change much from the prior mean. As the prior mean for μ increased, the posterior mean for τ^2 changed drastically from the prior mean for τ^2. The posterior density for ϕ becomes bimodal when the prior mean and prior variance for μ equals 10 and 3, respectively.

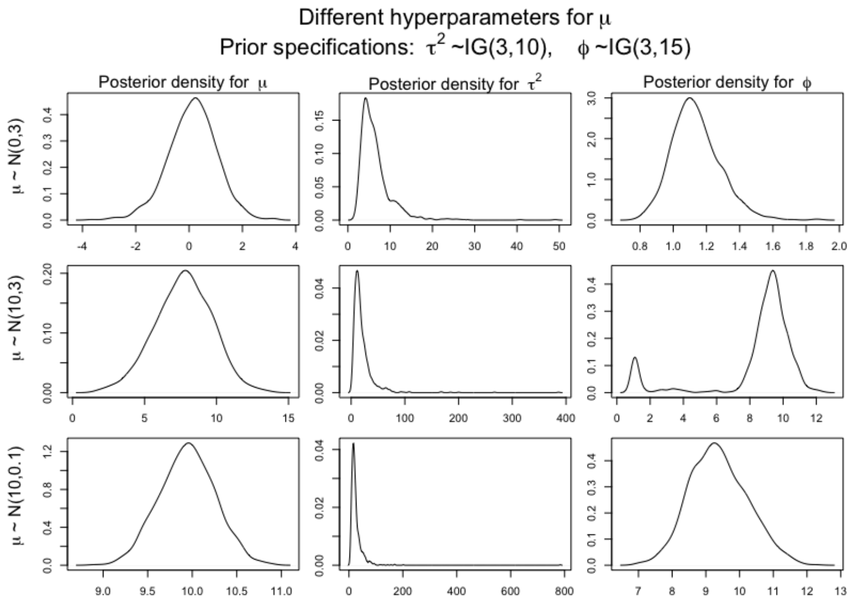

Figure 1: Density plots for the posterior samples of μ, τ^2 amd ϕ using different choices of hyperparameters for μ.

I repeated the process for Figure 2. In this case, the posterior means for μ don't change much. The posterior means for τ^2 matches the prior mean for τ^2.

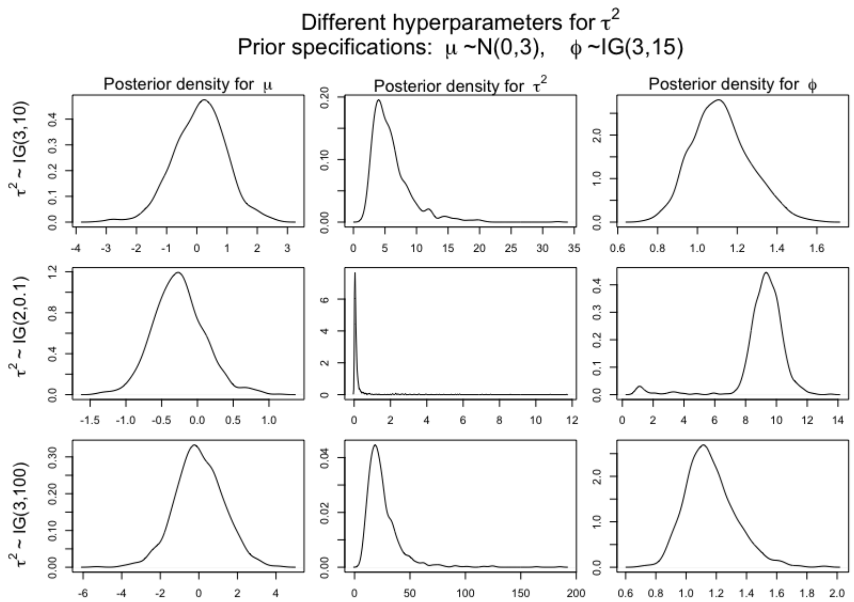

Figure 2: Density plots for the posterior samples of μ, τ^2 amd ϕ using different choices of hyperparameters for τ^2.

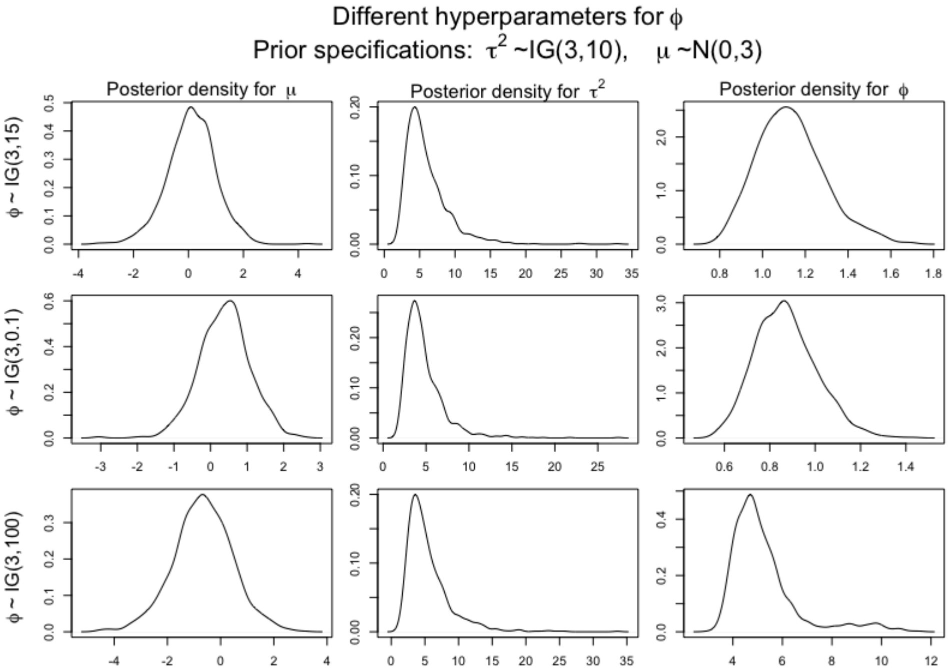

Figure 3: Density plots for the posterior samples of μ, τ^2 amd ϕ using different choices of hyperparameters for ϕ.

After examining the posterior predictive density (see 2.5), I decide ultimately to use the following prior specifications

$$\begin{aligned} \phi &\sim IG(3, 15) \\ \tau^2 &\sim IG(3, 10) \\ \mu &\sim N(0, 3) \end{aligned}$$

(2.3) **Prior specifications for $\boldsymbol{\alpha}$**. I chose two prior specifications for α to get an idea how it affects n^* and posterior predictive inference. I use $\alpha \sim Gamma(1, 1)$ and $\alpha \sim Gamma(2, .1)$. Based on figure 4, n^* is clearly affected by the choice of α. For $\alpha \sim Gamma(1, 1)$, n^* ranges from 5 to 9 and for $\alpha \sim Gamma(2, 0.1)$, n^* ranges from 8 to 29.

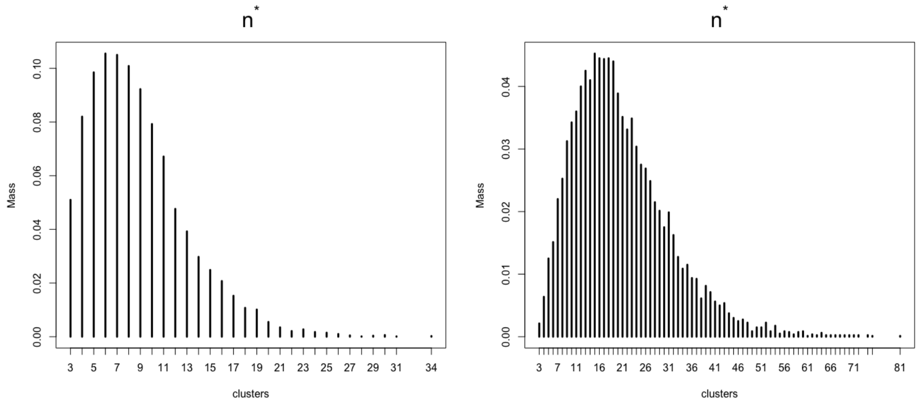

Figure 4: (Left) Posterior for n^* with prior $\alpha \sim Gamma(1, 1)$. (Right) Posterior for n^* with prior $\alpha \sim Gamma(2, 0.1)$.

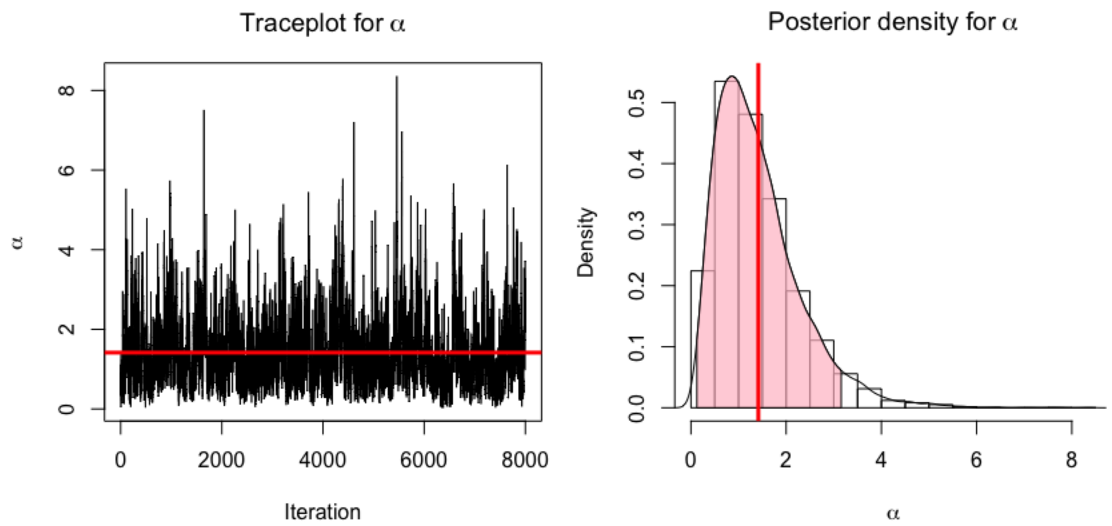

Figure 5: The traceplot and histogram for the posterior samples of α corresponding to a prior specified by $\alpha \sim Gamma(1, 1)$. The pink shaded region represents the 95% posterior credible interval for the parameter and the solid red line represents the posterior mean.

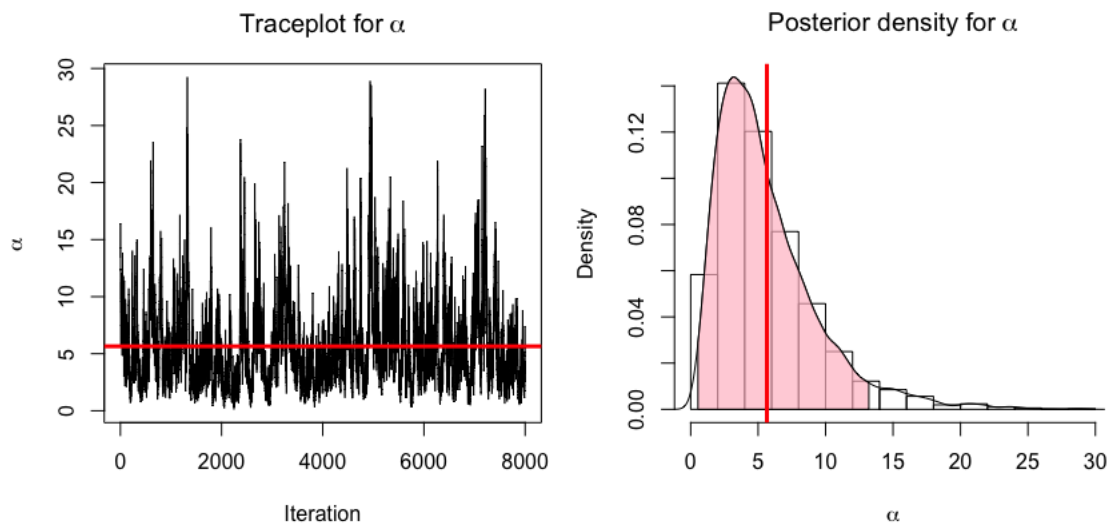

Figure 6: The traceplot and histogram for the posterior samples of α corresponding to a prior specified by $\alpha \sim Gamma(2, 0.1)$. The pink shaded region represents the 95% posterior credible interval for the parameter and the solid red line represents the posterior mean.

Based on figures 4 - 6, α and n^* are highly sensitive to the prior on α.

(2.4) **Clustering induced by the DP.** Figure 7 shows each cluster label θ_i, sorted by the order of the observations.

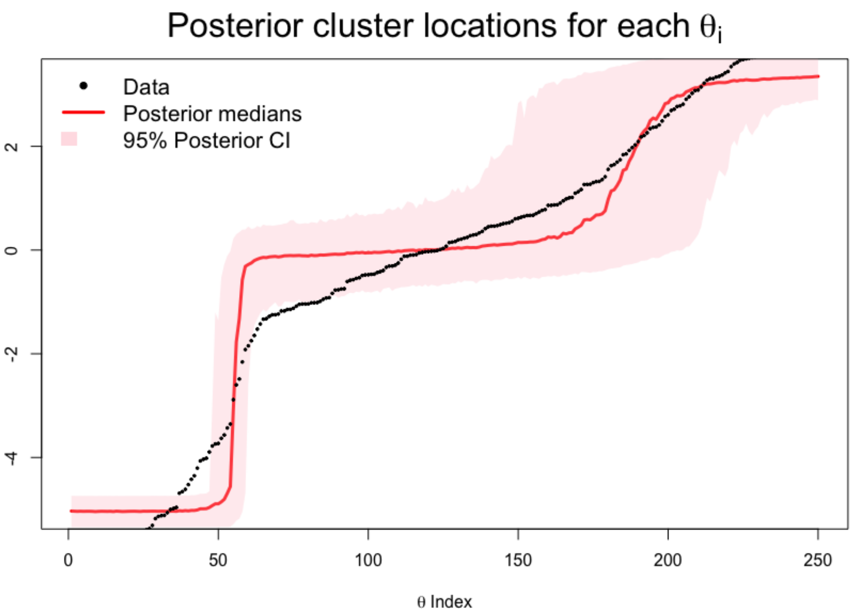

Figure 7: Posterior cluster locations for each θ_i sorted by the order of the magnitude of y.

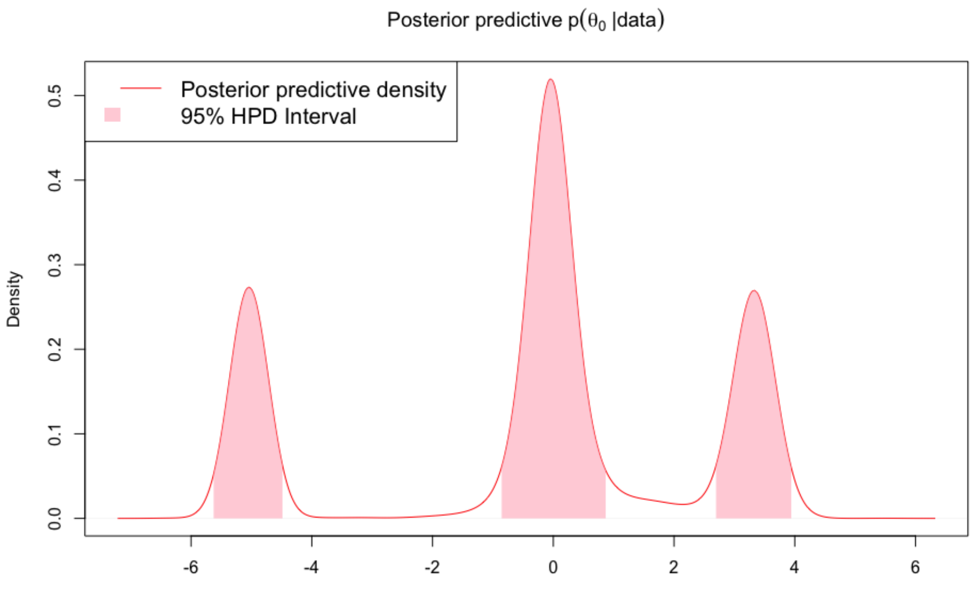

Figure 8: Posterior predictive for θ_0 with corresponding 95% highest posterior density interval.

Based on the above figures, there appears to be three clusters for θ's with medians at -5, 0, and 3. This is also consistent with the true model (i.e. The synthetic data generated from the mixture $0.2N(-5, 1) + 0.5N(0, 1) + 0.3N(3.5, 1)$).

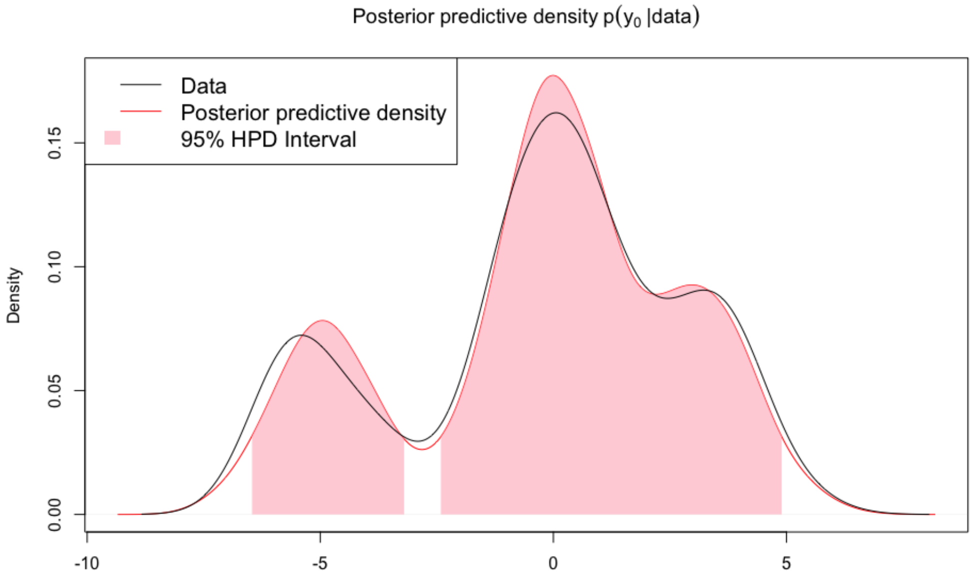

Figure 9: The posterior predictive for a new data point y_0, with corresponding highest posterior density interval shaded in pink.

(2.5) **Posterior predictive density $p(y_0|data)$**. Lastly, we obtain the posterior predictive density and use it to show that the prior specifications capture the distributional shape suggested by the data. Based on figure 9, the model was able to capture the underlying distribution of the data.

3. PROBLEM 3