Bayesian analysis using a Three-Component Mixture Model

Cynthia Kineza

2017-11-28

This year I took a course on Bayesian analysis. Many concepts were introduced such as Monte Carlo approximation, Markov Chain Monte Carlo / Gibb's Sampler,

I would suggest reading these books [Bayesian Methods for Data Analysis] (https://books.google.com/books/about/Bayesian_Methods_for_Data_Analysis_Third.html?id=GTJUt8fcFx8C), [A first course in Bayesian Statistical Methods] (https://www.jstor.org/stable/23116951?seq=1#page_scan_tab_contents)

In Chapters 10,11&12 in "Bayesian Methods for Data analysis", Bayesian computational concepts are introduced. I would suggest learning those concepts.

The project I was part of was about constructing a three component mixture model using Children's Behavorial Questionnaire(CBQ)

**CBQ: 195 items, assesses 15 aspects of temperament of children ages 3-8

***Low intensity pleasure: describes amount of pleasure or enjoyment related to situations involving low stimulus intensity, complexity, novelty Implications for self regulation, adolescent physical activity, etc. Not known if there are different groupings of children's LOWIP scores based on personality, parenting, gender, etc.

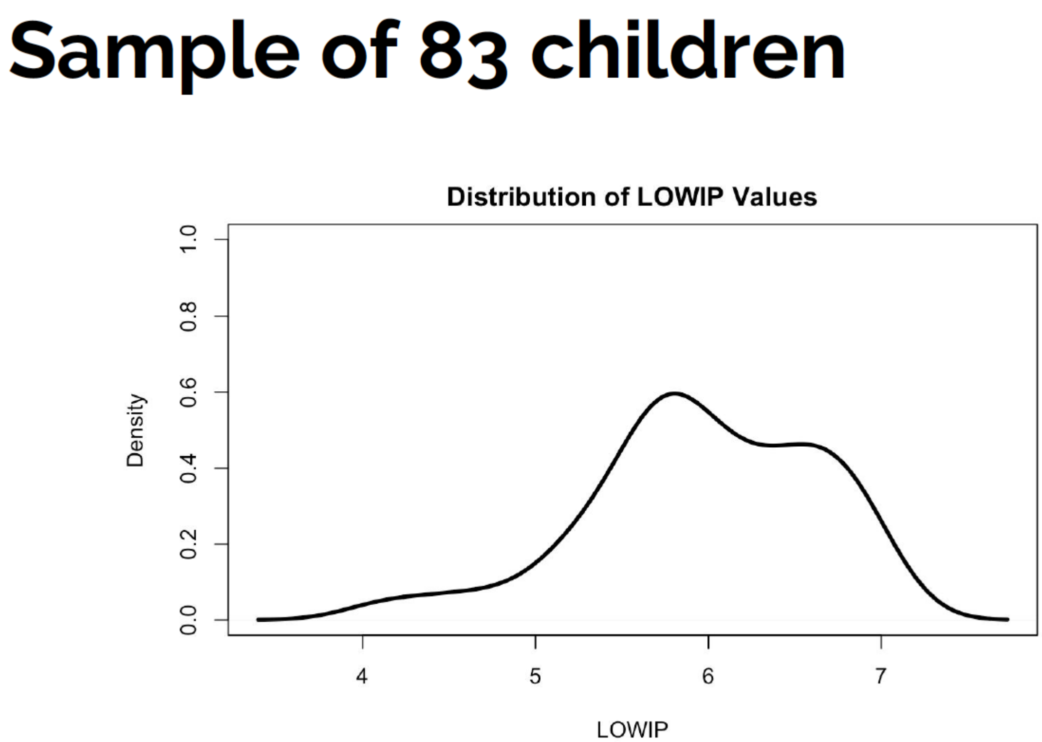

Figure 1: Caption for the picture.

Hypothesis

- ***DC:*** measure from Parental Stress Index describing how difficult or easy the mother perceives her child

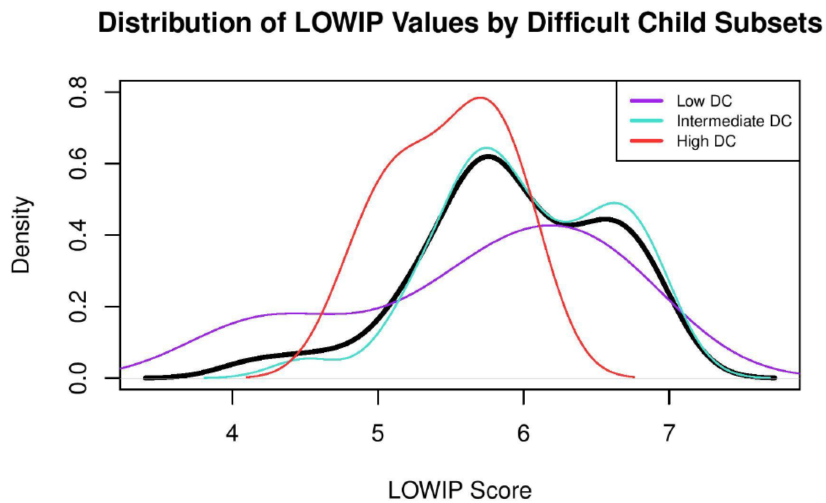

Figure 2: Caption for the picture.

Model

$$p(\theta_1|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\sigma}, p) = \text{normal}(\mu_{n_1}, \tau^2_{n_1}) \text{ where}$$

$$\tilde{\tau}^2_{n_1} = \tilde{\tau}^2_0 + n_1\tilde{\sigma}^2_1 \text{ and}$$

$$\mu_{n_1} = \frac{\tilde{\tau}^2_0}{\tilde{\tau}^2_0 + n_1\tilde{\sigma}^2_1} + \frac{n_1\tilde{\sigma}^2_1}{\tilde{\tau}^2_0 + n_1\tilde{\sigma}^2_1}\bar{y}_1,$$

$$\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i}$$

Figure 3: Caption for the picture.

Model

n_1 is the size of $\mathbf{Y}_1 = y_i : z_i = 1$. n_2 is the size of $\mathbf{Y}_2 = y_i : z_i = 2$. n_3 is the size of $\mathbf{Y}_3 = y_i : z_i = 3$.
The resulting full conditional $p|\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{y}, \boldsymbol{z}$ is $\text{rdirichlet}(1, c(n_1, n_2, n_3))$

Figure 4: Caption for the picture.

Model

$$z_i | y_i, \theta, \sigma^2, p \sim rmultinom\ (1, 1, c\ (\frac{(p_1 dnorm(y_i, \theta_1, \sigma_1^2)}{p_1 dnorm(y_i, \theta_1, \sigma_1^2) + p_2 dnorm(y_i, \theta_2, \sigma_2^2) + p_3 dnorm(y_i, \theta_3, \sigma_3^2)}, \frac{(p_2 dnorm(y_i, \theta_2, \sigma_2^2)}{p_1 dnorm(y_i, \theta_1, \sigma_1^2) + p_2 dnorm(y_i, \theta_2, \sigma_2^2) + p_3 dnorm(y_i, \theta_3, \sigma_3^2)}, \frac{(p_3 dnorm(y_i, \theta_3, \sigma_3^2)}{p_1 dnorm(y_i, \theta_1, \sigma_1^2) + p_2 dnorm(y_i, \theta_2, \sigma_2^2) + p_3 dnorm(y_i, \theta_3, \sigma_3^2)})$$

Figure 5: Caption for the picture.

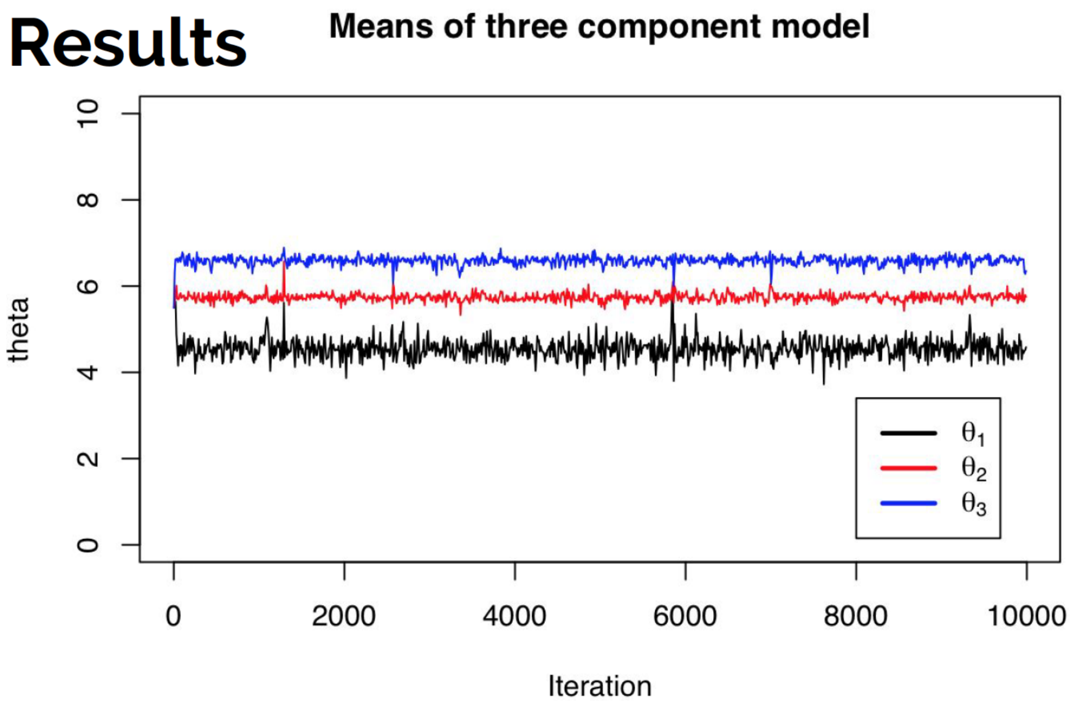

Figure 6: Caption for the picture.

Effective Size

Of 10,000 MCMC iterations:

- Component 1 - 947
- Component 2 - 873
- Component 3 - 819

Figure 7: Caption for the picture.

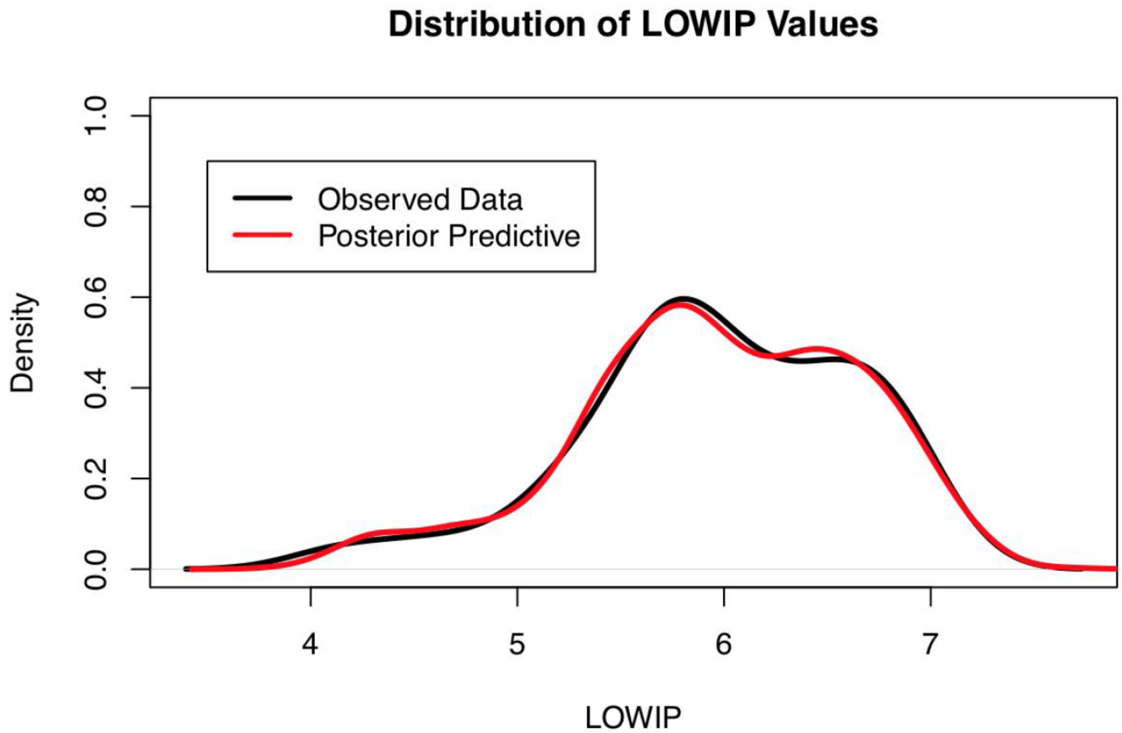

Figure 8: Caption for the picture.

Here, LOWIP was modeled well using a three-component model. Of the variables the research team provided, neither number of siblings or type of school attended seemed to be a good indicator.

Further research was found to be needed to explore what latent variables are responsible for this distribution of LOWIP Scores.

I hope this helped you to understand some concepts of bayesian analysis. The code of this project is included in my Github page.