STAT 631 Project
Gibbs Sampler for the (Extended) Marginal Rasch Model

Chenhui Hao

Abstract

In this project, a Markov chain -Monte Carlo method for Bayesian inference for the marginal Rasch model is constructed based on a theoretical important characterization of the marginal Rasch model showed by Cressie and Holland. This approach is developed and the operating characteristics are illustrated with simulated response data. Moreover, some important properties of marginal Rasch model are included. The estimates results I got are similar to the the ones in the reference paper[1].

Key words: Extended Rasch model, Gibbs sampler, item response models

1 Introduction

For item response theory (IRT) models, Markov chain - Monte Carlo (MCMC) approaches of Bayesian inference are very popular. The data augmentation Gibbs (DA-Gibbs) approach of Albert (1992)[2] is being followed by most of the applications. Data augmentation provides the technique where sampling from distributions is tractable. However, in the Markov chain both of the autocorrelation and computational cost of every step limit its usefulness for the large-scale applications.

In this project, an approach without using data augmentation is proposed. The Markov chain with lower autocorrelation is obtained as a result, at the mean time the tractable full conditional distributions are produced.

Started with the characterization of the marginal Rasch model by Cressie and Holland (1983)[3]. They show that, in their representation of the marginal Rasch model, only a limited number of characteristics of the ability distribution can be estimated without further assumptions on the distribution of ability. To solve this, based on Dutch identity (Holland, 1990)[4], we develop a parametrization of the marginal Rasch model in terms of item difficulty and Expected A Posteriori (EAP) estimators for ability, so that the examinees EAP estimators can be the marginal Rasch model even though their ability parameters can not. Also, this paper provides a Bayesian approach to statistical inference suitable for large-scale educational measurement contexts.

This paper is organized as follows. In section 2, the marginal Rasch model is introduced. In section 3, a Gibbs sampler of the marginal Rasch model (Cressie and Holland 1983[3]) is proposed. Section 4 provides the simulation and the results. Section 5 concludes my main contribution of the reproduction.

2 The (Extended) Marginal Rasch Model

The marginal Rasch model is expressed as follows:

$$P(\mathbf{X}=\mathbf{x}) = \int_{-\infty}^{\infty} \prod_i \frac{\exp\left(x_i\left(\theta-\delta_i\right)\right)}{1+\exp\left(\theta-\delta_i\right)} f(\theta)\mathrm{d}\theta \tag{1}$$

where function f denotes the density for the ability distribution, x_i is the binary response, $x_i = 1$ is correct answer and $x_i = 0$ refer to incorrect response, δ_i is the difficulty of item i, and θ is the ability. The following term can be written as

$$f(\theta|\mathbf{X}=\mathbf{0}) \propto f(\mathbf{X}=\mathbf{0}|\theta)f(\theta) = \frac{1}{\prod_i 1+\exp\left(\theta-\delta_i\right)} f(\theta)$$

It is proportional to the posterior distribution to the ability of some examinee that answers all items incorrectly i.e. $P0$. The marginal Rasch model can be expressed as follows

$$\begin{aligned} P(\mathbf{X}=\mathbf{x}) &= \int_{-\infty}^{\infty} \prod_i \exp\left(x_i\left(\theta-\delta_i\right)\right) f(\theta|\mathbf{X}=\mathbf{0})\mathrm{d}\theta P(\mathbf{0}) \\ &= \int_{-\infty}^{\infty} \exp\left(\sum_i x_i\left(\theta-\delta_i\right)\right) f(\theta|\mathbf{X}=\mathbf{0})\mathrm{d}\theta P(\mathbf{0}) \\ &= \left(\prod_i b_i^{x_i}\right) \mathcal{E}\left(\exp\left(x_+\Theta\right)|\mathbf{X}=\mathbf{0}\right) P(\mathbf{0}) \end{aligned} \tag{2}$$

where $b_i = \exp\left(-\delta_i\right)$ and x_+ denotes the sum score. From equation 2, it is clear that the full population distribution can not be derived from the marginal Rasch model. However, the equation 2 can be treated as a characterization of the marginal Rasch model which is useful to construct a Gibbs sampler.
With the change of some notations

$$\begin{aligned} \mu &= P(\mathbf{0}) \\ \lambda_s &= \mathcal{E}(\exp(s\Theta)|\mathbf{X}=\mathbf{0}) \end{aligned}$$

The characterization of the marginal Rasch model is obtained:

$$P(\mathbf{x}) = \prod b_i^{x_i}\lambda_{x+}\mu \tag{3}$$

In order to ensure that equation 3 represents a probability distribution i.e. $\sum_x P(\mathbf{x}) = 1$, a constraint is applied as follows

$$\begin{aligned} \sum_x P(\mathbf{x}) &= \sum_x \prod_i b_i^{x_i}\lambda_{x+}\mu = 1 \\ \mu &= \frac{1}{\sum_{\mathrm{x}} \prod_i b_i^{x_i}\lambda_{x+}} = \frac{1}{\sum_{s=0}^{n} \gamma_s(\mathbf{b})\lambda_s} \end{aligned} \tag{4}$$

Function γ_s denotes the elementary symmetric function of order s of the vector $\mathbf{b}$

$$\gamma_s(\mathbf{b}) = \sum_{\mathbf{x}\rightarrow s} \prod_i b_i^{x_i}$$

where the sum goes through all response patterns x that yield the sum score s.
Substitute the constraint equation 4 in to equation 3, the marginal Rasch model becomes to following expression.

$$\begin{aligned} P(\mathbf{x}) &= \prod b_i^{x_i} \lambda_{x_+} \mu \\ &= \frac{\prod_i b_i^{x_i} \lambda_{x_+}}{\sum_{s=0}^{n} \gamma_s(\mathbf{b}) \lambda_s} \\ &= p(\mathbf{x}|\mathbf{b}, \lambda) \end{aligned} \tag{5}$$

It is apparent that this marginal Rasch model represents a probability distribution for all values of its parameters.
Some additional structures can be derived from properties of the marginal Rasch model. Therefore, we are going to focus on some properties that are important both theoretically and practically.
First, the following factorization is found from equation 5

$$\begin{aligned} p(\mathbf{x}|\mathbf{b}, \lambda) &= p\left(\mathbf{x}|x_+, \mathbf{b}\right) p\left(x_+|\mathbf{b}, \lambda\right) \\ &= \frac{\prod_i b_i^{x_i}}{\gamma_{x_+}(\mathbf{b})} \frac{\gamma_{x_+}(\mathbf{b}) \lambda_{x_+}}{\sum_{s=0}^{n} \gamma_s(\mathbf{b}) \lambda_s} \\ &= \frac{\prod_i b_i^{x_i}}{\gamma_{x_+}(\mathbf{b})} \pi_{x_+} \end{aligned} \tag{6}$$

The factorization shows that the observed score distribution is statistic for λ, notice that parameters $\mathbf{b}, \lambda$ and $\mathbf{b}, \pi$ are one-one transformations of each other. The last expression of equation 6 is called the extended Rasch model by Cressie and Holland (1983)[3].
Second, the marginal and conditional distributions corresponding to euqation 5 are being considered. Specifically, the distribution of $\mathbf{x}$ without item n ($\mathbf{x}^{(n)}$) can be derived by

$$\begin{aligned} p\left(\mathbf{x}^{(n)}|\mathbf{b}, \lambda\right) &= p\left(\mathbf{x}^{(n)}, 1|\mathbf{b}, \lambda\right) + p\left(\mathbf{x}^{(n)}, 0|\mathbf{b}, \lambda\right) \\ &= \frac{\prod_{i \neq n} b_i^{x_i} \left(\lambda_{x_+^{(n)}} + \lambda_{x_+^{(n)}+1} b_n\right)}{\sum_{s=0}^{n} \gamma_s(\mathbf{b}) \lambda_s} \\ &= \frac{\prod_{i \neq n} b_i^{x_i} \left(\lambda_{x_+^{(n)}} + \lambda_{x_+^{(n)}+1} b_n\right)}{\sum_{s=0}^{n} \left(\gamma_s\left(\mathbf{b}^{(n)}\right) + \gamma_{s-1}\left(\mathbf{b}^{(n)}\right) b_n\right) \lambda_s} \\ &= \frac{\prod_{i \neq n} b_i^{x_i} \left(\lambda_{x_+^{(n)}} + \lambda_{x_+^{(n)}+1} b_n\right)}{\sum_{s=0}^{n-1} \gamma_s\left(\mathbf{b}^{(n)}\right) \left(\lambda_s + \lambda_{s+1} b_n\right)} \end{aligned} \tag{7}$$

where the last two expressions are according to the following recursive property of elementary symmetric functions (Verhelst, Glas, van der Sluis, 1984)[5]:

$$\gamma_s(\mathbf{b}) = \gamma_s\left(\mathbf{b}^{(i)}\right) + \gamma_{s-1}\left(\mathbf{b}^{(i)}\right) b_i \tag{8}$$

Therefore, it shows that $\mathbf{X}^{(n)}$ is also a marginal Rasch model. The distribution of X_n conditioning on the remaining $n-1$ responses is obtained.

$$\begin{aligned} p\left(X_n = x|\mathbf{x}^{(n)}, \mathbf{b}, \lambda\right) &= \frac{b_n^x \lambda_{x_+^{(n)}+x}}{\eta_{x_+^{(n)}}} \\ &= \frac{\left(b_n \frac{\lambda_{x_+^{(n)}+1}}{\lambda_{x_+^{(n)}}}\right)^x}{1 + b_n \frac{\lambda_{x_+^{(n)}+1}}{\lambda_{x_+^{(n)}}}} = p\left(X_n = x|x_+^{(n)}, \mathbf{b}, \lambda\right) \end{aligned} \tag{9}$$

It is showing that this conditional distribution only depends on the remaining n 1 responses by the raw score x_+^i, and it is independent to the remaining item parameters $\mathbf{b}^{(n)}$. Third, besides changing the parameterizatio, one constraint is needed, it is that the model of equation 3 can reduce to model of equation 2 if and only if the parameters λ_S represent a sequence of moments. To illustrate the constraint more explicitly, consider λ_1 and λ_2. Since

$$\mathrm{Var}(X) = E\left(X^2\right) - (E(X))^2$$

The following expression is obtained

$$\lambda_2 = \mathcal{E}(\exp(2\Theta)|\mathbf{X}=\mathbf{0}) \geq \mathcal{E}(\exp(\Theta)|\mathbf{X}=\mathbf{0})^2 = \lambda_1^2$$

Therefore, the inequality constraints could be formulated as follows

$$\det\begin{bmatrix} \lambda_0 & \lambda_1 & \dots & \lambda_m \\ \lambda_1 & \lambda_2 & \dots & \lambda_{m+1} \\ \vdots & \vdots & & \vdots \\ \lambda_m & \lambda_{m+1} & \dots & \lambda_{2m} \end{bmatrix} \geq 0, \text{and} \det\begin{bmatrix} \lambda_1 & \lambda_2 & \dots \\ \lambda_2 & \lambda_3 & \dots \\ \vdots & \vdots & \vdots \\ \lambda_{m+1} & \lambda_{m+2} & \dots\lambda_{2m+1} \end{bmatrix} \geq 0$$

where $m = 0, 1, 2, \ldots$ (Shohat Tamarkin, 1943)[6].
In the next section, after a Gibbs sampler for the extended Rasch model in equation 3 is introduced, the additional constraints implied in equation 2 are incorporated in the method. Fourth, the parameters λ_s are difficult to interpret even if all the moment constraints are satisfied, because λ_s represent a sequence of moments corresponding to the expected posterior distribution of ability for a person that answers all items incorrectly. According to this, a new parametrization is introduced from the Dutch identity (Holland, 1990)[4]. After applied to the marginal Rasch model, the following is obtained

$$\begin{aligned} \tau_s = \frac{\lambda_{s+1}}{\lambda_s} &= \frac{\mathcal{E}(\exp((s+1)\Theta)|\mathbf{X}=\mathbf{0})}{\mathcal{E}(\exp((s)\Theta)|\mathbf{X}=\mathbf{0})} \\ &= \frac{\int_{-\infty}^{\infty}\exp((s+1)\theta)f(\theta|\mathbf{X}=\mathbf{0})\mathrm{d}\theta}{\int_{-\infty}^{\infty}\exp(s\theta)f(\theta|\mathbf{X}=\mathbf{0})\mathrm{d}\theta} \\ &= \int_{-\infty}^{\infty}\exp(\theta)\frac{\frac{\exp(s\theta)}{\prod_i 1+\exp(\theta-\delta_i)}f(\theta)}{\int_{-\infty}^{\infty}\frac{\exp(s\theta)}{\prod_i 1+\exp(\theta-\delta_i)}f(\theta)\mathrm{d}\theta}\mathrm{d}\theta \\ &= \int_{-\infty}^{\infty}\exp(\theta)f\left(\theta|X_+=s\right)\mathrm{d}\theta = \mathcal{E}\left(\exp(\Theta)|X_+=s\right) \end{aligned} \tag{10}$$

The last expression is the posterior expectation of ability for different scores. This new parametrization is useful when the moment constraints are implied in the marginal Rasch model. With this parameterization, the marginal Rasch model can be expressed as follows in terms of the item parameters $\mathbf{b}$ and the EAP parameters $\boldsymbol{\tau}$.

$$P(\mathbf{X}=\mathbf{x}|\mathbf{b},\boldsymbol{\tau}) = \frac{\prod_i b_i^{x_i}\prod_{s<x_+}\tau_s}{\sum_s\gamma_s(\mathbf{b})\prod_{t<s}\tau_t}$$

Fifth, a further property of the Dutch identity is that not only EAP estimators for ability can be obtained, but more generally, the following expression can be obtained as well.

$$\frac{\lambda_{s+t}}{\lambda_s} = \mathcal{E}\left(\exp(t\Theta)|X_+=s\right) \quad, 0 \leq s+t \leq n \tag{11}$$

In order to show that this can be used to sample from the parameters posterior distribution to obtain estimates of both the posterior mean and variance of ability, we obtain the following expressions using equation 11.

$$\mathcal{E}\left(\exp(\Theta)|X_+ = s, \mathbf{b}, \lambda\right) = \frac{\lambda_{s+1}}{\lambda_s}, s = 0, \ldots, n-1$$

and

$$\mathcal{E}\left(\exp(\Theta)^2|X_+ = s, \mathbf{b}, \lambda\right) = \mathcal{E}\left(\exp(2\Theta)|X_+ = s, \mathbf{b}, \lambda\right) = \frac{\lambda_{s+2}}{\lambda_s}, s = 0, \ldots, n-2$$

Then the following can be derived.

$$\begin{aligned}\mathcal{E}\left(\exp(\Theta)|X_+ = s, \mathbf{X} = \mathbf{x}\right) &= \mathcal{E}\left[\mathcal{E}\left(\exp(\Theta)|X_+ = s, \mathbf{B}, \mathbf{\Lambda}\right)|\mathbf{X} = \mathbf{x}\right] \\ &= \mathcal{E}\Big[\frac{\Lambda_{s+1}}{\Lambda_s}|\mathbf{X} = \mathbf{x}\Big], \quad s = 0, \ldots, n-1\end{aligned}$$

This can be directly estimated by Monte Carlo integration with the sample from the posterior distribution of $\mathbf{\Lambda}$. Similarly, the posterior variance of ability can be estimated as

$$\begin{aligned}\mathcal{V}\left(\exp(\Theta)|X_+ = s, \mathbf{X} = \mathbf{x}\right) =& \mathcal{V}\left[\mathcal{E}\left(\exp(\Theta)|X_+ = s, \mathbf{B}, \mathbf{\Lambda}\right)|\mathbf{X} = \mathbf{x}\right] \\ &+ \mathcal{E}\left[\mathcal{V}\left(\exp(\Theta)|X_+ = s, \mathbf{B}, \mathbf{\Lambda}\right)|\mathbf{X} = \mathbf{x}\right]\end{aligned} \tag{12}$$

for $s = 0, \ldots, n-1$, and $\mathcal{V}\left(\exp(\Theta)|X_+ = s, \mathbf{b}, \lambda\right)$ is estimated as follows:

$$\begin{aligned}\mathcal{V}\left(\exp(\Theta)|X_+ = s, \mathbf{b}, \lambda\right) &= \mathcal{E}\left(\exp(2\Theta)|X_+ = s, \mathbf{b}, \lambda\right) - \mathcal{E}^2\left(\exp(\Theta)|X_+ = s, \mathbf{b}, \lambda\right) \\ &= \frac{\lambda_{s+2}}{\lambda_s} - \left(\frac{\lambda_{s+1}}{\lambda_s}\right)^2\end{aligned}$$

In equation 12, the first part of right hand side is uncertain because of the unknown parameters $\mathbf{b}$ and λ, while the second part is uncertain due to the finite test length. Specifically, the first part of equation tends to zero when the number of examinees goes to infinity. The second part goes to zero when the number of items tends to infinity.

Finally, the purpose of the study in the area of educational surveys is to relate ability to examinees characteristics. The research based on the marginal Rasch model is introduced briefly below. In general applications, the relationship between persons responses ($\mathbf{Y}$) and characteristics ($\mathbf{X}$), such as gender, are independent conditioning on the ability. Generally, the distribution of ability conditioning on $\mathbf{Y}$ is modelled as a normal regression model.

Theorem 1. If $\mathbf{Y} \perp \mathbf{X}|\Theta$ and $\mathbf{X} \perp \Theta|X_+$, then also $\mathbf{Y} \perp \mathbf{X}|X_+$.
Proof. The following expression of joint distribution is implied to prove the theorem.

$$f\left(\mathbf{x}, x_+, \mathbf{y}, \theta\right) = f(\mathbf{y}|\theta)p\left(\mathbf{x}|x_+\right)p\left(x_+|\theta\right)f(\Theta)$$

Then the following equation can be obtained.

$$f\left(\mathbf{y}, \mathbf{x}|x_+\right) = p\left(\mathbf{x}|x_+\right)\int_{-\infty}^{\infty} f(\mathbf{y}|\theta)f\left(\theta|x_+\right)\mathrm{d}\theta. \quad \square$$

Theorem 1 shows that under the assumptions of independence between $\mathbf{Y}$, $\mathbf{X}$ conditioning on θ, the information between $\mathbf{Y}$ and $\mathbf{X}$ is contained in the distribution of $\mathbf{Y}$ conditioning on $\mathbf{X}_+$, Notice that Theorem 1 holds true for every element of $\mathbf{Y}$. This means that the main effects of people characteristics with the item relating to an element of $\mathbf{Y}$ and the rest to $\mathbf{X}_+$ can be considered. Additionally, the distribution of X_+ conditioning on an element from $\mathbf{Y}$ can be well estimated by Bayesian analysis.

3 Gibbs Sampler

In order to sample from the posterior distribution of $\boldsymbol{b}$ and $\boldsymbol{\lambda}$ corresponding to equation 5, a prior distribution needs to be specified. We consider a prior that can derive the tractable full conditional distributions for each parameter. This prior is given as:

$$f(\mathbf{b}, \lambda) = \Big(\prod_i \alpha_i b_i^{\alpha_i - 1}\Big)\Big(\prod_s \beta_s \lambda_s^{\beta_s - 1}\Big) \tag{13}$$

We assume that no item is answered correctly or incorrectly by all examinees, and that every score occurs at least once, all α_i and β_s are equal to one. By using this prior, the posterior distribution can be derived as:

$$\begin{aligned} f(\boldsymbol{b}, \boldsymbol{\lambda}|\boldsymbol{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}) &\propto f(\boldsymbol{x}|\boldsymbol{b}, \boldsymbol{\lambda}) f(\boldsymbol{b}, \boldsymbol{\lambda}|\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ &= \frac{\prod_i b_i^{x_i} \lambda_{x_+}}{\sum_{s=0}^n \gamma_s(\mathbf{b})\lambda_s} \prod_i \alpha_i b_i^{\alpha_i - 1} \prod_s \beta_s \lambda_s^{\beta_s - 1} \\ &\propto \frac{\prod_i b_i^{x_{+i} + \alpha_i - 1} \prod_s \lambda_s^{m_s + \beta_s - 1}}{(\sum_{s=0}^n \gamma_s(\mathbf{b})\lambda_s)^m} \end{aligned} \tag{14}$$

where x_{+i} is the number of persons that answer item i correctly, m_s refers to the number of the number of persons that obtain the sum score s and m denotes the number of people. The distribution of equation 14 is not one of the standard form distributions. Therefore, We implement a Gibbs Sampler , so that we can obtain full conditional distributions that are easy to sample from and generate a Markov chain for which the posterior distribution in equation 14 is the invariant distribution.

3.1 Full Conditional Distribution for b_i

The full conditional distribution for parameter b_i is propportional to

$$f\left(b_i|\mathbf{b}^{(i)}, \lambda, \mathbf{x}; \alpha, \beta\right) \propto \frac{b_i^{x_{+i} + \alpha_i - 1}}{(\sum_{s=0}^n \gamma_s(\mathbf{b})\lambda_s)^m} \tag{15}$$

In order to generate samples from the full conditional distribution in equation 15, we use the recursive property of elementary symmetric functions in equation. 8. The result of equation 8 shows that elementary symmetric functions are linear in each of their arguments, so we substitute equation 8 and rewrite the full conditional distribution of euqation 15.

$$\begin{aligned} f\left(b_i|\mathbf{b}^{(i)}, \lambda, \mathbf{x}; \alpha, \beta\right) &\propto \frac{b_i^{x_{+i} + \alpha_i - 1}}{(\sum_{s=0}^n \gamma_s(\mathbf{b})\lambda_s)^m} \\ &= \frac{b_i^{x_{+i} + \alpha_i - 1}}{(\sum_{s=0}^n (\gamma_s(\mathbf{b}^{(i)}) + \gamma_{s-1}(\mathbf{b}^{(i)}) b_i)\lambda_s)^m} \\ &= \frac{b_i^{x_{+i} + \alpha_i - 1}}{(\sum_{s=0}^n \gamma_s(\mathbf{b}^{(i)})\lambda_s + \sum_{s=0}^n \gamma_{s-1}(\mathbf{b}^{(i)}) b_i \lambda_s)^m} \\ &= \frac{b_i^{x_{+i} + \alpha_i - 1}}{(\sum_{s=0}^n \gamma_s(\mathbf{b}^{(i)})\lambda_s)^m \left(1 + \frac{\sum_{s=0}^n \gamma_{s-1}(\mathbf{b}^{(i)}) b_i \lambda_s}{\sum_{s=0}^n \gamma_s(\mathbf{b}^{(i)})\lambda_s}\right)^m} \\ &\propto \frac{b_i^{x_{+i} + \alpha_i - 1}}{(1 + c b_i)^m} \end{aligned} \tag{16}$$

where c is constant only depending on other parameters,

$$c = \frac{\sum_{s=0}^{n} \gamma_{s-1}\left(\mathbf{b}^{(i)}\right)\lambda_s}{\sum_{s=0}^{n} \gamma_s\left(\mathbf{b}^{(i)}\right)\lambda_s}.$$

With the variables transformation,

$$y = \frac{cb_i}{1+cb_i} \rightarrow b_i = \frac{y}{c(1-y)} \tag{17}$$

The following expression is obtained

$$\begin{aligned} f\left(y|\mathbf{b}^{(i)}, \lambda, \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}\right) &\propto \frac{\left(\frac{y}{c(1-y)}\right)^{x_{+i}+\alpha_i-1}}{\left(\frac{1}{1-y}\right)^m} \\ &= \frac{y^{x_{+i}+\alpha_i-1}}{(1-y)^{x_{+i}+\alpha_i-1}(1-y)^{-m}} \\ &= y^{x_{+i}+\alpha_i-1}(1-y)^{m-x_{+i}-\alpha_i+2-1} \end{aligned} \tag{18}$$

We can see this is a beta distribution, $\mathrm{Beta}\left(x_{+i}+\alpha_i, m-x_{+i}-\alpha_i+2\right)$. Therefore, if we generate y from beta distribution, b_i can be obtained correspondingly from the full conditional distribution in equation 15. Formally, the distribution in equation. 15 is classified as a scaled Beta prime distribution.

3.2 Full Conditional Distribution for λ_s

The full conditional distribution for parameter λ is proportional to

$$f\left(\lambda_t|\mathbf{b}, \lambda^{(t)}, \mathbf{x}; \alpha, \boldsymbol{\beta}\right) \propto \frac{\lambda_t^{m_t+\beta_t-1}}{\left(\sum_{s=0}^{n} \gamma_s(\mathbf{b})\lambda_s\right)^m} \tag{19}$$

When considering the full conditional distribution for the item parameters, we found that the denominator of equation 19 is linear in λ_t, such that

$$\begin{aligned} f\left(\lambda_t|\mathbf{b}, \lambda^{(t)}, \mathbf{x}; \alpha, \boldsymbol{\beta}\right) &\propto \frac{\lambda_t^{m_t+\beta_t-1}}{\left(\sum_{s=0}^{n} \gamma_s(\mathbf{b})\lambda_s\right)^m} \\ &= \frac{\lambda_t^{m_t+\beta_t-1}}{\left(\sum_{s\neq t} \gamma_s(\mathbf{b})\lambda_s + \gamma_t(\mathbf{b})\lambda_t\right)^m} \\ &= \frac{\lambda_t^{m_t+\beta_t-1}}{\left(\sum_{s\neq t} \gamma_s(\mathbf{b})\lambda_s\right)^m \left(1+\frac{\gamma_t(\mathbf{b})\lambda_t}{\sum_{s\neq t}\gamma_s(\mathbf{b})\lambda_s}\right)^m} \propto \frac{\lambda_t^{m_t+\beta_t-1}}{(1+c\lambda_t)^m} \end{aligned} \tag{20}$$

where the constant c is

$$c = \frac{\gamma_t(\mathbf{b})}{\sum_{s\neq t} \gamma_s(\mathbf{b})\lambda_s}$$

With the same parameter transformation,

$$z = \frac{c\lambda_t}{1+c\lambda_t} \rightarrow \lambda_t = \frac{z}{c(1-z)}$$

The following expression is obtained

$$f\left(z|\lambda^t, \mathbf{b}, \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}\right) \propto z^{m_t+\beta_t-1}(1-z)^{m-m_t-\beta_t+2-1} \tag{21}$$

From the result, We see that the full conditional distributions for both parameters b_i and λ_t follow the same family of Beta distributions.

4 Simulation in R

In this section, the Gibbs Sampler is implemented in R to do the simulations, then simulation results are presented. In the simulation, there are 30 items, the items difficulties follow a standard uniform distribution, the abilities of 100,000 examinees are generated from a standard normal distribution. Since the data is very large matrix with 3000,000 elements, the computational cost is big, I only run 1000 iterations, from the results, the Markov chain of the method converge within the iteration times. The initial values for $\mathbf{b}$ and λ are both uniformly distributed from 0 to 1.

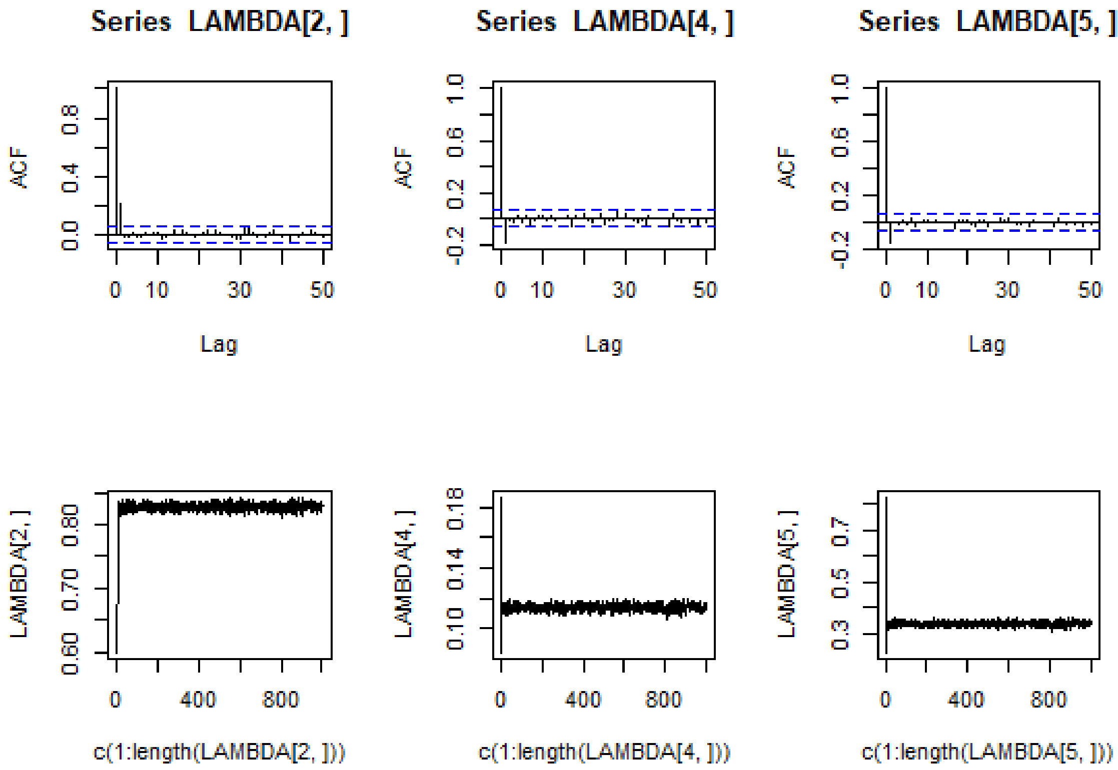

Figure 1: ACF plot and trace plot for $\lambda_6, \lambda_7, \lambda_8$

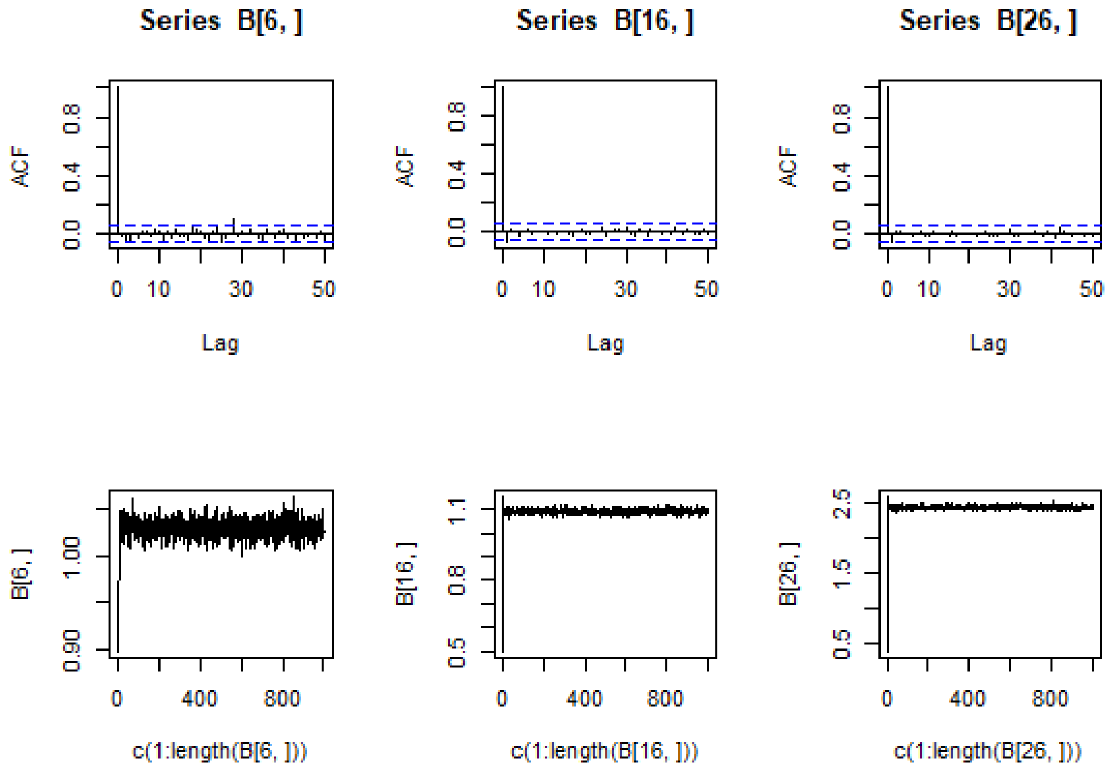

Figure 2: ACF plot and trace plot for b_6, b_{16}, b_{26}

To check the convergence in Markov chain, I evaluate the autocorrelation which is a function of lag, and convergence of the Gibbs sampler of the Markov chain. Figure 1 shows the autocorrelation and the trace plot for the parameters $\lambda_2, \lambda_4, \lambda_5$, Figure 2 shows the autocorrelation and the trace plot for the parameters b_6, b_{16}, b_{26}. From the autocorrelation function and the trend of the trace plot, we can tell that the chain of this Gibbs sampler converges. In conclude, the Markov chain generates an independent and identical distributed samples from the full conditional distribution with low autocorrelation.

5 Conclusion

In this project, I reproduced the derivation of the model and programmed in R with reference to the illustrative code provided by the paper[1], I also modified some mistakes in their derivation and programmings in the paper and got the similar results from original reference paper[1]. Based on the paper, I derived the marginal Rasch model, considered some significant properties of the model including the conditional distribution of x without the item, the moment constraint applied in the model and the Expected A Posterior estimators. The purpose of this paper is to learn the properties of the marginal Rasch model and how to illustrate the operating characteristics of marginal Rasch model by an efficient Bayesian inference approach.

References

[1] A Gibbs Sampler for the (Extended) Marginal Rasch Model Maris, Gunter ; Bechger, Timo ; Martin, Ernesto Psychometrika, 2015, Vol.80(4), pp.859-879

[2] Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. Journal of Educational Statistics, 17(3), 251–269.

[3] Cressie, N., Holland, P. (1983). Characterizing the manifest probabilities of latent trait models. Psychometrika, 48, 129–141.

[4] Holland, P. W. (1990). The dutch identity: A new tool for the study of item response models. Psychometrika, 55, 5–18.

[5] Verhelst, N. D., Glas, C. A.W., van der Sluis, A. (1984). Estimation problems in the Rasch model: The basic symmetric functions. Computational Statistics Quarterly, 1(3), 245–262.

[6] Shohat, J. H., Tamarkin, J. D. (1943). The problem of moments. New York: American Mathematics Society.*Soil Science Society of America Journal* 60, 433-442.