

Bayesian Q Estimation with Attribute Hierarchy

Heesang Ann

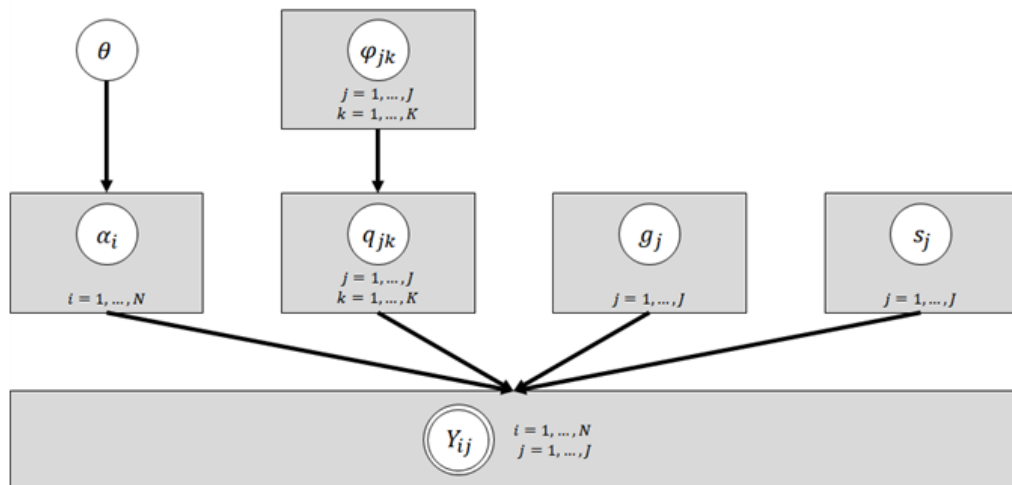
Bayesian Models for Q

- Suggested by Chung(2014) & Chen(2018) → Same Gibbs Sampling method Essentially!
- Variables for Estimation

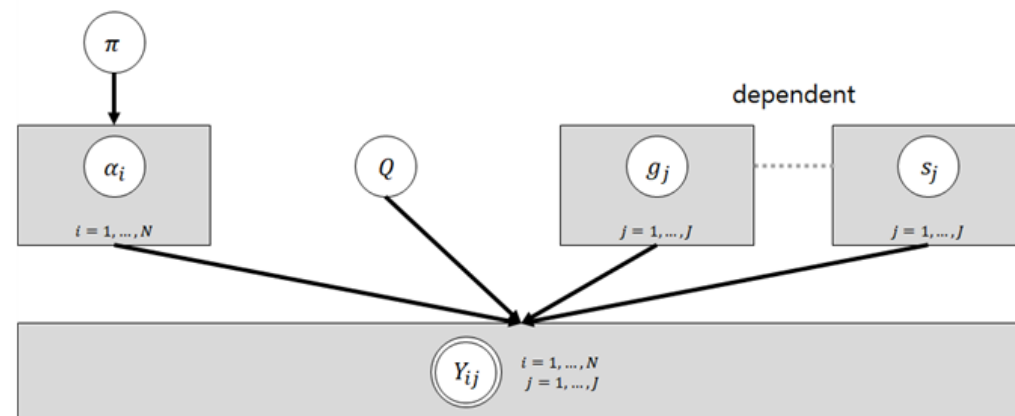
Variables	Notation	Prior
Attribute pattern	$\alpha_i, \text{ for } i = 1, \dots, N$	For $c \in \{0, 1, \dots, 2^K - 1\}$, $\Pr(\alpha_i = a_c) = \pi_c$
Class distribution	$\pi = (\pi_0, \dots, \pi_{2^K - 1})$	$\pi \sim \text{dirichlet}(1, 1, \dots, 1)$
Item parameter	$g_j \text{ and } s_j, \text{ for } j = 1, \dots, J$	$g_j \sim \text{Beta}(\alpha_g, \beta_g)$ and $s_j \sim \text{Beta}(\alpha_s, \beta_s)$, with constraint $0 \leq g_j < 1 - s_j \leq 1$ $p(s_j, g_j) \propto s_j^{\alpha_s - 1} (1 - s_j)^{\beta_s - 1} g_j^{\alpha_g - 1} (1 - g_j)^{\beta_g - 1} I(0 \leq g_j < 1 - s_j \leq 1)$
Q matrix	$Q = (q_{jk})_{J \times K}$	Chung) $\Pr(q_{jk} = 1) = p_{jk}, p_{jk} \sim \text{Beta}(a, b)$
($\Pr(q_{jk} = 1)$)	$p_{jk}, \text{ for } j = 1, \dots, J, k = 1, \dots, K$	Chen) Uniform on \mathbb{Q} (Identifiability Q Set)

Bayesian Models for Q

Bayesian Model (Chung, 2014)



Bayesian Model (Chen, 2018)



Likelihood on DINA Model

$$P(Y_{ij} = 1 | \alpha_i, q_j, s_j, g_j) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}} \quad , \quad \text{where} \quad \eta_{ij} = \prod_k \alpha_{ik}^{q_{ik}}$$

$$\text{fully,} \quad P(Y | \alpha_i, q_j, s_j, g_j) = \prod_{i=1}^N \prod_{j=1}^J ((1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}})^{y_{ij}} (s_j^{\eta_{ij}} (1 - g_j)^{1 - \eta_{ij}})^{1 - y_{ij}}$$

Gibbs Sampling Process

[Step1] Update Attribute pattern

Prior(pmf) : $\Pr(\alpha_i = a_c | \pi) = \pi_c$ for each class $c = 0, 1, \dots, 2^K - 1$

Likelihood : $P(Y_i = \mathbf{y}_i | \alpha_i = a_c, Q, s, g) = \prod_{j=1}^J [(1-s_j)^{\eta_{c,j}} g_j^{1-\eta_{c,j}}]^{y_{ij}} [s_j^{\eta_{c,j}} (1-g_j)^{1-\eta_{c,j}}]^{1-y_{ij}}, \quad \eta_{c,j} = \prod_{k=1}^K a_{c(k)}^{q_k}$

\Rightarrow Posterior(pmf) : Prior x Likelihood

- Use the fact 'log(prior) + log(likelihood) = log(posterior)'
- Use appropriate matrices to update all attribute patterns at the same time

Gibbs Sampling Process

[Step2] Update Class Distribution

Prior : $\pi \sim \text{dirichlet}(1,1,\dots,1)$

Likelihood : $P(\alpha_1, \dots, \alpha_N | \pi) = \pi_0^{n_0} \pi_1^{n_1} \dots \pi_{2^K-1}^{n_{2^K-1}}$ where n_c : # of attribute pattern a_c

\Rightarrow Posterior : $\pi | \alpha \sim \text{dirichlet}(1+n_0, 1+n_1, \dots, 1+n_{2^K-1})$

- Dirichlet distribution is the conjugate prior of Multinomial distribution!

Gibbs Sampling Process

[Step3] Update Item Parameter (Chen)

$$\text{Prior : } p(s_j, g_j) \propto s_j^{\alpha_s - 1} (1 - s_j)^{\beta_s - 1} g_j^{\alpha_g - 1} (1 - g_j)^{\beta_g - 1} I(0 \leq g_j < 1 - s_j \leq 1)$$

$$\text{Likelihood : } P(y_{1j}, \dots, y_{Nj} | \alpha, Q, s_j, g_j) = \prod_{i=1}^N [(1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}}]^{y_{ij}} [s_j^{\eta_{ij}} (1 - g_j)^{1 - \eta_{ij}}]^{1 - y_{ij}}$$

$$\Rightarrow \text{Posterior : } p(s_j, g_j | y_{1j}, \dots, y_{Nj}, \alpha, Q) \propto s_j^{\tilde{\alpha}_s - 1} (1 - s_j)^{\tilde{\beta}_s - 1} g_j^{\tilde{\alpha}_g - 1} (1 - g_j)^{\tilde{\beta}_g - 1} I(0 \leq g_j < 1 - s_j \leq 1)$$

$$\text{where } \tilde{\alpha}_s = \sum_{i=1}^N (1 - y_{ij}) \eta_{ij} + \alpha_s, \quad \tilde{\beta}_s = \sum_{i=1}^N y_{ij} \eta_{ij} + \beta_s, \quad \tilde{\alpha}_g = \sum_{i=1}^N y_{ij} (1 - \eta_{ij}) + \alpha_g, \quad \tilde{\beta}_g = \sum_{i=1}^N (1 - y_{ij}) (1 - \eta_{ij}) + \beta_g$$

- Chung(2014) used the fact that Beta distribution is the conjugate prior of Binomial distribution, ad conjugate prior and a quantile function for inverse transform to maintain $0 \leq g_j < 1 - s_j \leq 1$
 → Consequently, the same result!

Gibbs Sampling Process

[Step4] Update Q-matrix (Chung)

$$\text{Prior : } \Pr(\mathbf{q}_j = \mathbf{a}_c | p_{j1}, \dots, p_{jk}) \propto \prod_{k=1}^K [p_{jk}^{a_{c(k)}} (1 - p_{jk})^{1 - a_{c(k)}}] \text{ and } \Pr(\mathbf{q}_j = (0, 0, \dots, 0)) = 0.$$

$$\text{Likelihood : } P(y_{1j}, \dots, y_{Nj} | \mathbf{q}_j = \mathbf{a}_c, \boldsymbol{\alpha}, s_j, g_j) = \prod_{i=1}^N [(1 - s_j)^{\eta_{ij}^{(c)}} g_j^{1 - \eta_{ij}^{(c)}}]^{y_{ij}} [s_j^{\eta_{ij}^{(c)}} (1 - g_j)^{1 - \eta_{ij}^{(c)}}]^{1 - y_{ij}}$$

Posterior(pmf) : Prior x Likelihood

- Update q-vector of Q-matrix one by one.
- Use the fact 'log(prior) + log(likelihood) = log(posterior)'

Gibbs Sampling Process

[Step5] Update Probability for $q=1$ (Chung)

Prior : $p_{jk} \sim \text{Beta}(a, a)$

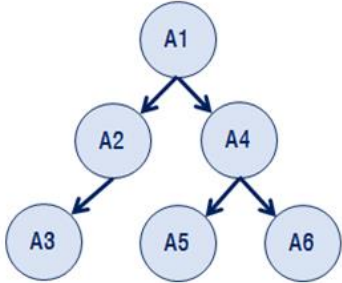
Likelihood : $P(q_{jk} | p_{jk}) = \begin{cases} p_{jk} & \text{for } q_{jk} = 1 \\ 1 - p_{jk} & \text{for } q_{jk} = 0 \end{cases}$

\Rightarrow Posterior : $p_{jk} | q_{jk} \sim \text{Beta}(a + q_{jk}, a + 1 - q_{jk})$

- Beta distribution is the conjugate prior of Binomial distribution!

Attribute Hierarchy

- Words : preceding attribute, following attribute, root attribute, adjacent matrix, access matrix, possible attribute pattern
- Expressions for the hierarchy information :

Direct Graph	Adjacent Matrix	Access matrix
 <pre> graph TD A1((A1)) --> A2((A2)) A1 --> A4((A4)) A2 --> A3((A3)) A4 --> A5((A5)) A4 --> A6((A6)) </pre>	$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$R = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

Develop a more efficient and accurate estimation method for Q-matrix with the hierarchy information

Bayesian Model - H

- Hierarchy structure -> fixed meaning of attributes
- A : adjacent matrix for hierarchy information.
- QI : Expert suggested matrix

Each entries of QI means the probability of '(corresponding q-entry)=1

Examples of QI matrix

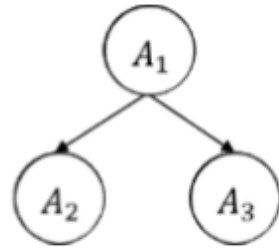
	A1	A2	A3
P1	1	0.3	0.3
P2	0.7	1	0.3
P3	0.7	0.3	0.7
P4	0.3	0.7	0.3
P5	0.3	0.7	0.7
:	:	:	:

	A1	A2	A3
P1	0.8	0.3	0.1
P2	0.9	0.7	0.2
P3	0.8	0.2	0.7
P4	0.1	0.9	0.4
P5	0.2	0.8	0.7
:	:	:	:

- λ : Prior Effect size (default $\lambda = 1$)

Bayesian Model – Strong H

[Possible Attribute Patterns]



$a_0 = (0,0,0)$	$c_1 = 0$
$a_1 = (0,0,1)$	X
$a_2 = (0,1,0)$	X
$a_3 = (0,1,1)$	X
$a_4 = (1,0,0)$	$c_2 = 4$
$a_5 = (1,0,1)$	$c_3 = 5$
$a_6 = (1,1,0)$	$c_4 = 6$
$a_7 = (1,1,1)$	$c_5 = 7$

[Priors for Possible Attribute Pattern & Restricted Class Distribution]

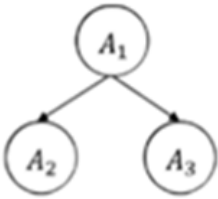
Variables	Notation	Prior
Attribute pattern	$\alpha_i, \text{ for } i = 1, \dots, N$	For each classnum $c_l \in C_H, \Pr(\alpha_i = a_{c_l}) = \pi_{c_l}$, and for $c \in \{0, 1, \dots, 2^K - 1\} - C_H, \Pr(\alpha_i = a_c) = 0$
Restricted Class distribution	$\pi = (\pi_{c_1}, \dots, \pi_{c_L})$	$\pi \sim \text{dirichlet}(1, 1, \dots, 1)$

- In the situation of strong hierarchy, prior for attribute patterns other than possible attribute patterns set 0.

Bayesian Model - Strong H

[The equivalence relationship of q-vector]

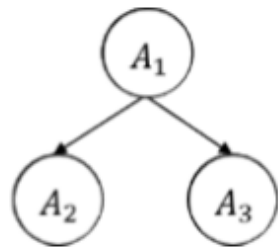
- In the situation of strong hierarchy, two different q-vectors can have the same likelihood
- Define the equivalence relationship if two different q-vectors have the same likelihood

Hierarchy	Att. Pattern	$q_j = (0,1,0)$	$q_j = (1,1,0)$
	(0,0,0)	$\eta_{ij} = 0$	$\eta_{ij} = 0$
	(1,0,0)	$\eta_{ij} = 0$	$\eta_{ij} = 0$
	(1,0,1)	$\eta_{ij} = 0$	$\eta_{ij} = 0$
	(1,1,0)	$\eta_{ij} = 1$	$\eta_{ij} = 1$
	(1,1,1)	$\eta_{ij} = 1$	$\eta_{ij} = 1$

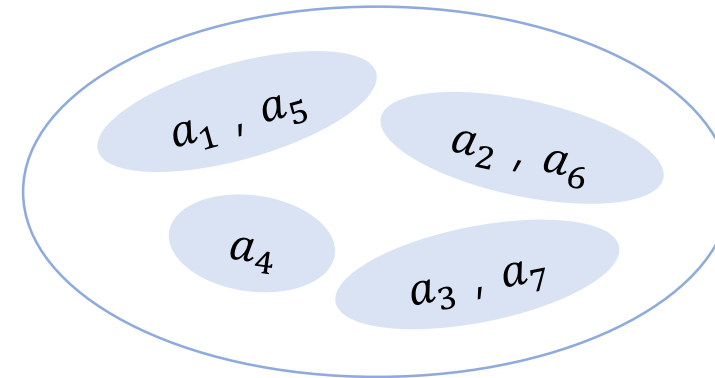
→ Equivalence Class : $\{(1,0,0)\}$, $\{(0,1,0), (1,1,0)\}$, $\{(0,0,1), (1,0,1)\}$, $\{(0,1,1), (1,1,1)\}$

Bayesian Model - Strong H

- (# of equivalence classes for q-vector) = (# of possible attribute patterns) - 1
 \therefore Each classes have 1-1 correspondence with a non-zero possible attribute pattern.
- Notation for the equivalence classes : $\widetilde{a_{c_2}}, \widetilde{a_{c_3}}, \dots, \widetilde{a_{c_i}}$



$a_0 = (0,0,0)$	$c_1 = 0$
$a_1 = (0,0,1)$	X
$a_2 = (0,1,0)$	X
$a_3 = (0,1,1)$	X
$a_4 = (1,0,0)$	$c_2 = 4$
$a_5 = (1,0,1)$	$c_3 = 5$
$a_6 = (1,1,0)$	$c_4 = 6$
$a_7 = (1,1,1)$	$c_5 = 7$



For the efficiency, first, determine the equivalence classes for each q-vectors, then, decide q-vector by the closest q-vector form the experts' suggestion.

Bayesian Model - Strong H

- qnum: the equivalence class number for a q-vector (For q-vectors in \widetilde{a}_{c_l} , qnum = l)
- Qnum: the collection of qnum


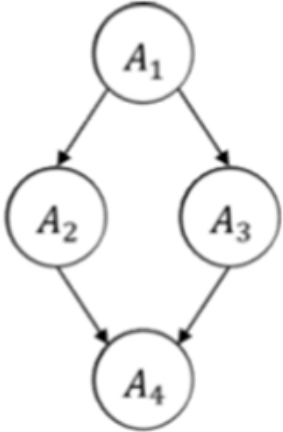
[Qnum과 qnum distribution의 Prior]

Variables	Notation	Prior
Qnum	$Qn = (qn_1, \dots, qn_J)^T$	For $c_l \in C_H - \{c_1\}$, $\Pr(qn_j = l \mid \phi) = \Pr(\mathbf{q}_j \in \widetilde{a}_{c_l}) = \phi_{c_l}$
qnum distribution	$\phi_j = (\phi_{c_2}^{(j)}, \dots, \phi_{c_L}^{(j)}),$ for $j = 1, \dots, J$	$\phi_j \sim \text{dirichlet}(\lambda p_{c_2}^{(j)}, \dots, \lambda p_{c_L}^{(j)}), \text{for } p_{c_l}^{(j)} = \sum_{a_c \in \widetilde{a}_{c_l}} P(\mathbf{q}_j = a_c) / (1 - p_0^{(j)})$ Each $P(\mathbf{q}_j = a_c)$ and $p_0^{(j)} = P(\mathbf{q}_j = (0, 0, \dots, 0))$ can be computed by QI matrix. (Defaul : $\lambda = 1$)

Data Simulation

3. Experiment

- $N = 1000$ (same with Chung(2014))
- Hierarchy structure & true Q matrices (used by Chen et al.(2016))

Structure1 ($K=3, J=10$)	Structure2 ($K=4, J=18$)																																																																																																																																											
<div><pre>graph TD; A1((A1)) --> A2((A2)); A2 --> A3((A3));</pre></div> <table><tr><th></th><th>A_1</th><th>A_2</th><th>A_3</th></tr><tr><td>I1</td><td>1</td><td>0</td><td>0</td></tr><tr><td>I2</td><td>0</td><td>1</td><td>0</td></tr><tr><td>I3</td><td>0</td><td>0</td><td>1</td></tr><tr><td>I4</td><td>1</td><td>0</td><td>0</td></tr><tr><td>I5</td><td>0</td><td>1</td><td>0</td></tr><tr><td>I6</td><td>0</td><td>0</td><td>1</td></tr><tr><td>I7</td><td>1</td><td>1</td><td>0</td></tr><tr><td>I8</td><td>1</td><td>0</td><td>1</td></tr><tr><td>I9</td><td>0</td><td>1</td><td>1</td></tr><tr><td>I10</td><td>1</td><td>1</td><td>1</td></tr></table>		A_1	A_2	A_3	I1	1	0	0	I2	0	1	0	I3	0	0	1	I4	1	0	0	I5	0	1	0	I6	0	0	1	I7	1	1	0	I8	1	0	1	I9	0	1	1	I10	1	1	1	<div><pre>graph TD; A1((A1)) --> A2((A2)); A1 --> A3((A3)); A2 --> A4((A4)); A3 --> A4;</pre></div> <table><tr><th></th><th>A_1</th><th>A_2</th><th>A_3</th><th>A_4</th></tr><tr><td>I1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>I2</td><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>I3</td><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>I4</td><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>I5</td><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>I6</td><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>I7</td><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>I8</td><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>I9</td><td>1</td><td>1</td><td>0</td><td>0</td></tr><tr><td>I10</td><td>1</td><td>0</td><td>1</td><td>0</td></tr><tr><td>I11</td><td>1</td><td>0</td><td>0</td><td>1</td></tr><tr><td>I12</td><td>0</td><td>1</td><td>1</td><td>0</td></tr><tr><td>I13</td><td>0</td><td>1</td><td>0</td><td>1</td></tr><tr><td>I14</td><td>0</td><td>0</td><td>1</td><td>1</td></tr><tr><td>I15</td><td>1</td><td>1</td><td>1</td><td>0</td></tr><tr><td>I16</td><td>1</td><td>1</td><td>0</td><td>1</td></tr><tr><td>I17</td><td>1</td><td>0</td><td>1</td><td>1</td></tr><tr><td>I18</td><td>0</td><td>1</td><td>1</td><td>1</td></tr></table>		A_1	A_2	A_3	A_4	I1	1	0	0	0	I2	0	1	0	0	I3	0	0	1	0	I4	0	0	0	1	I5	1	0	0	0	I6	0	1	0	0	I7	0	0	1	0	I8	0	0	0	1	I9	1	1	0	0	I10	1	0	1	0	I11	1	0	0	1	I12	0	1	1	0	I13	0	1	0	1	I14	0	0	1	1	I15	1	1	1	0	I16	1	1	0	1	I17	1	0	1	1	I18	0	1	1	1
	A_1	A_2	A_3																																																																																																																																									
I1	1	0	0																																																																																																																																									
I2	0	1	0																																																																																																																																									
I3	0	0	1																																																																																																																																									
I4	1	0	0																																																																																																																																									
I5	0	1	0																																																																																																																																									
I6	0	0	1																																																																																																																																									
I7	1	1	0																																																																																																																																									
I8	1	0	1																																																																																																																																									
I9	0	1	1																																																																																																																																									
I10	1	1	1																																																																																																																																									
	A_1	A_2	A_3	A_4																																																																																																																																								
I1	1	0	0	0																																																																																																																																								
I2	0	1	0	0																																																																																																																																								
I3	0	0	1	0																																																																																																																																								
I4	0	0	0	1																																																																																																																																								
I5	1	0	0	0																																																																																																																																								
I6	0	1	0	0																																																																																																																																								
I7	0	0	1	0																																																																																																																																								
I8	0	0	0	1																																																																																																																																								
I9	1	1	0	0																																																																																																																																								
I10	1	0	1	0																																																																																																																																								
I11	1	0	0	1																																																																																																																																								
I12	0	1	1	0																																																																																																																																								
I13	0	1	0	1																																																																																																																																								
I14	0	0	1	1																																																																																																																																								
I15	1	1	1	0																																																																																																																																								
I16	1	1	0	1																																																																																																																																								
I17	1	0	1	1																																																																																																																																								
I18	0	1	1	1																																																																																																																																								

Data Simulation

- Master probability for root attributes & conditional master probability

	Structure1	Structure2
	$P(A_1 = 1) = r_1, P(A_1 = 0) = 1 - r_1,$ $P(A_2 = 1 A_1 = 1) = r_{2,1},$ $P(A_2 = 1 A_1 = 0) = r_{2,0}$ $P(A_3 = 1 A_2 = 1) = r_{3,1},$ $P(A_3 = 1 A_2 = 0) = r_{3,0}$	$P(A_1 = 1) = r_1, P(A_1 = 0) = 1 - r_1,$ $P(A_2 = 1 A_1 = 1) = r_{2,1}, P(A_2 = 1 A_1 = 0) = r_{2,0}$ $P(A_3 = 1 A_1 = 1) = r_{3,1}, P(A_3 = 1 A_1 = 0) = r_{3,0}$ $P(A_4 = 1 A_2A_3 = 1) = r_{4,1}, P(A_4 = 1 A_2A_3 = 0) = r_{4,0}$
Default	$r_1 = 0.7, r_{2,1} = 0.8, r_{3,1} = 0.8$ Strong H : $r_{2,0} = r_{3,0} = 0$ Weak H : $r_{2,0}, r_{3,0} \sim Unif[0,0.1]$	$r_1 = 0.7, r_{2,1} = 0.8, r_{3,1} = 0.8, r_{4,1} = 0.8$ Strong H : $r_{2,0} = r_{3,0} = r_{4,0} = 0$ Weak H : $r_{2,0}, r_{3,0}, r_{4,0} \sim Unif[0,0.1]$

- Default for item parameters : $s_j = g_j = 0.2$ (same with Chung(2014))

Future Work

- Accuracy of Bayesian Estimation - SH
- Suggest appropriate λ
- Develop Bayesian Model - Weak Hierarchy
- Compare previous methods, in terms of accuracy and efficiency
- Studies about the effects by various factors
- ...