Generating matrices

with ordinal responses and fixed margins

Kathrin Gruber & Reinhold Hatzinger

Department of Statistics and Mathematics, WU Wien

The Rasch Model (Rasch, 1960)

$$P(X_{vi} = 1|\theta_v, \beta_i) = \frac{exp(\theta_v - \beta_i)}{1 + exp(\theta_v - \beta_i)}$$

0	0	1	0	2
1	0	0	1	2
1	0	0	0	1
1	0	1	1	3
0	0	1	0	1
1	1	1	0	3
0	1	0	0	1
1	0	0	0	1
1	1	0	0	2
0	0	0	1	1
6	3	4	3	

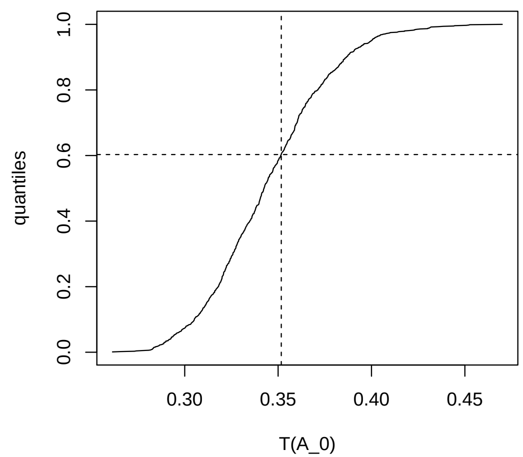
- completely determined by the margins
- model fit can be evaluated by parametric and quasi-exact tests

Motivation for exact tests

Statistical tests and confidence intervals are based on exact probability statements that are valid for any sample size.

Construction principle:

- Rearrange the labels of the observed data points.
- Calculate all possible values of the test statistic.
- Yields the distribution of the test statistic under the null hypothesis.



Motivation for exact tests (cont.)

Advantages:

- No parameter estimation necessary.
- Are not based on asymptotic and approximate statistical methods.
- Valid for small sample sizes.

Background

Various algorithms for sampling 0–1 matrices with given marginals in a nonuniform way have already been proposed and can generally be divided into two classes:

- Nonuniform sampling schemes:
  - Recursive solving of a linear program with restrictions to the row sums.
  - Based on the sequential importance sampling (SIS) algorithm (e.g. Snijders, 1991, Chen and Small, 2005, Chen, Dinwood and Sullivant, 2006).

1	.	.	2
0	.	.	1
1	.	.	1
2	1	1	

→

1	0	1	2
0	1	0	1
1	0	0	1
2	1	1	

Background (cont.)

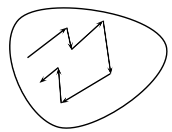
- Applications of the Markov-Chain Monte Carlo method:
  - All matrices in the sample space are considered as states.
  - The sampling scheme and a special permutation rule is defining their transition probability (Ponocny, 2001, Verhelst, 2008).

0	1
1	0
1	0
1	1
0	0

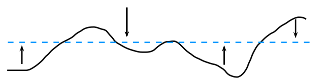
→

1	0
0	1
1	0
1	1
0	0

Requirements



- Coverage of the whole sample space
- Independence



- Uniform sampling

The MCMC approach of Verhelst in general

- All binary matrices with fixed row (r) and column (c) sums, $A \in \Sigma_{rc}$ (whereas Σ_{rc} denotes the sample space of possible matrices) are considered as states.
- The observed data matrix is considered as the starting state, A_0.
- A_0 can be transformed in one step into other matrices $A_t \in \Sigma_{rc}$ using a well defined rule R.
- The R-neighborhood is the set of all reachable matrices using such a transformation of A_0, $\mathcal{A}_R(A_0)$.
- Sampling algorithm:

$$\boxed{A_0} \xrightarrow[\mathcal{A}_R(A_0)]{} \boxed{A_1} \xrightarrow[\mathcal{A}_R(A_1)]{} \boxed{A_2} \cdots\cdots \boxed{A_s} \xrightarrow[\mathcal{A}_R(A_s)]{\cdots\cdots} \boxed{A_{s+1}}$$

- Sampling scheme defines the transition matrix $P = (p_{st})$ (with $\lim_{n\to\infty} P^n e_t = \pi$) of the Markov Chain

The binomial rule and binomial neighborhoods

Binomial rule:

→ Assign a ones to m rows to the first column with row totals equal to one, and zero to the $m - a$ rows.

→ Yields first column of the transformed matrix, the second one is just the compliment of it.

1	2	3	4
0	0	1	0
1	0	0	1
1	0	0	0
0	1	1	1
0	0	1	0
1	1	1	0
0	1	0	0
1	0	0	0
1	1	0	0
0	0	0	1

→

1	2
0	0
1	0
1	0
0	1
0	0
1	1
0	1
1	0
1	1
0	0

→

1	2
0	0
0	1
1	0
1	0
0	0
1	1
1	0
0	1
1	1
0	0

→

1	2	3	4
0	0	1	0
0	1	0	1
1	0	0	0
1	0	1	1
0	0	1	0
1	1	1	0
1	0	0	0
0	1	0	0
1	1	0	0
0	0	0	1

The binomial rule and binomial neighborhoods (cont.)

The B_{ij}-neighborhood of $A \in \Sigma_{rc}$ is defined by

$$\mathcal{A}_B^{(i,j)}(A) = \{A_s : A_s \text{ is a } B_{ij} \text{ transform of } A \text{ and } A_s \neq A\}$$

The set of all matrices that can be formed by a single binomial transformation of a single column pair of A is

$$\mathcal{A}_B(A) = \bigcup_{(i,j)} \mathcal{A}_B^{(i,j)}(A)$$

1	2	3	4
1	1	1	0
1	1	0	1
1	0	1	0
0	1	0	0
0	1	0	1
0	0	0	1
1	1	1	0

1	2
1	1
1	1
1	0
0	1
0	1
0	0
1	1

$\binom{3}{1}$

1×2

1	3
1	1
1	0
1	1
0	0
0	0
0	0
1	1

–

1×0

...

1	4
1	0
1	1
1	0
0	0
0	1
0	1
1	0

$\binom{3}{2}$

1×2

The binomial rule and binomial neighborhoods (cont.)

1	2	3	4
1	1	1	0
1	1	0	1
1	0	1	0
0	1	0	0
0	1	0	1
0	0	0	1
1	1	1	0

1	2
1	1
1	1
1	0
0	1
0	1
0	0
1	1

1	3
1	1
1	0
1	1
0	0
0	0
0	0
1	1

...

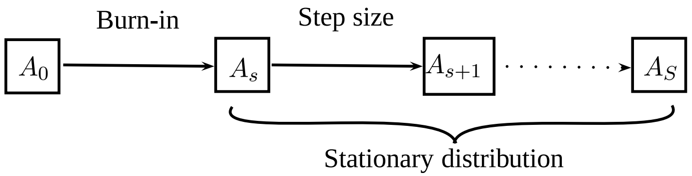
1	4
1	0
1	1
1	0
0	0
0	1
0	1
1	0

The column pair (i,j) is a Guttman pair if $a_{ij} \times b_{ij} = 0$, if $a_{ij} \times b_{ij} > 0$ the pair is called regular.

The k_2-measure of $A \in \Sigma_{rc}$ is defined as

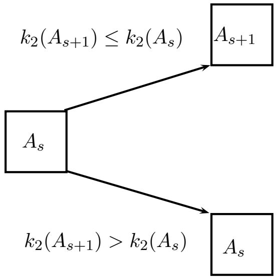
$$k_2(A) = \{\sharp(i,j) : i < j \leq k, (i,j) \text{ is a regular pair}\}$$

The MCMC-method in general



The Metropolis-Hastings algorithm

- Start the chain in the observed data matrix, A_0.
- Select randomly a pair of columns (i, j) from the $k_2(A_s)$ regular column pairs of A_s.
- Apply a random binomial operation to the selected pair, yielding A_{s+1}.
  - If $A_{s+1} = A_s$ repeat step 2.
  - Otherwise:



The Partial Credit Model (Masters, 1982)

A generalization of the probability for a response of person v on category h $(h = 0, \ldots, m_i)$ of item i, whereas m_i is the number of response categories.

$$P(X_{vih} = 1) = \frac{exp(h\theta_v + \beta_{ih})}{\sum_{l=0}^{m_i} exp(l\theta_v + \beta_{il})}$$

X_{vih} ... person v scores in category h of item i
θ_v ... location of person v on latent trait h
β_{ih} ... item category combination (allows different numbers of response categories)

Requirements

1. Coverage of the whole sample space
2. Independence
3. Uniform sampling
4. Frequency distribution for categories must be maintained for each item.

Binomial neighborhoods in ordinal data

Apply a binomial transformation on a single category tupel (g, h) of a defined set of integer tupels $C^{(i,j)}$ (= all possible category combinations).

1	2	3	4
0	2	1	0
1	2	0	1
1	0	2	0
2	0	1	1
0	2	1	0
2	0	0	2
2	2	0	0
0	0	1	0
1	0	0	2
0	0	0	1

→

1	2
0	2
1	2
1	0
2	0
0	2
2	0
2	2
0	0
1	0
0	0

→

1	2
2	0
1	2
1	0
0	2
2	0
0	2
2	2
0	0
1	0
0	0

→

1	2	3	4
2	0	1	0
1	2	0	1
1	0	2	0
0	2	1	1
2	0	1	0
0	2	0	2
2	2	0	0
0	0	1	0
1	0	0	2
0	0	0	1

Binomial neighborhoods in ordinal data (cont.)

The generalized B_{ij}-neighborhood of a matrix $B \in \Sigma_{rc}$ is defined as

$$\mathcal{A}_B(B) = \bigcup_{(i,j)} \left(\mathcal{A}_B{}^{(i,j)}(B) \times C^{(i,j)} \right)$$

B is a matrix with maximum x categories of integers in increasing order.
$C^{(i,j)}$ is the set of all category tupels $(g, h), g < h$ of a column pair $(i, j) \in B$.

(g,h) \ (i,j)	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
(0,1)	x		x			
(0,2)	x				x	
(1,2)	x	x				

Extension to multinomial neighborhoods

Apply the transformation for each category tupel $(g, h), g < h$, simultaneously.

1	2	3	4
0	2	1	0
1	2	0	1
1	0	2	0
2	0	1	1
0	2	1	0
2	0	0	2
2	2	0	0
0	0	1	0
1	0	0	2
0	0	0	1

→

1	2
0	2
1	2
1	0
2	0
0	2
2	0
2	2
0	0
1	0
0	0

→

1	2
2	0
2	1
1	0
0	2
2	0
0	2
2	2
0	0
0	1
0	0

→

1	2	3	4
2	0	1	0
2	1	0	1
1	0	2	0
0	2	1	1
2	0	1	0
0	2	0	2
2	2	0	0
0	0	1	0
0	1	0	2
0	0	0	1

Extension to multinomial neighborhoods (cont.)

The M_{ij}-neighborhood of matrix $B \in \Sigma_{rc}$ is defined as

$$\mathcal{B}_M(B) = \mathcal{P}\left(\mathcal{A}_{B_{(g,h)}}^{(i,j)}(B) \right)$$

Set of simultaneous binomial transforms on a single column pair (i, j) applied to each of its category tupels (g, h).
The multinomial neighborhood is the power set of the column pair subset D and the category tupel subset $C^{(i,j)}$.

$$\mathcal{A}_{B_{(g,h)}}^{(i,j)}(B) = D \times C^{(i,j)}$$

(g,h) \ (i,j)	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
(0,1)	x		x			
(0,2)	x				x	
(1,2)	x	x				

Outlook

- Investigate the behavior of the algorithm.
  - First rule for small data sizes.
  - Second rule for big data sizes.
- Check required...
  - Burn-in period (stationarity)
  - Rejection rates of the MH
  - Step size
- Develop quasi-exact tests for the family of partial credit models.

Y. Chen, P. Diaconis, S.P. Holmes, and J.S. Liu. Sequential Monte Carlo Methods for Statistical Analysis of Tables. *Journal of the American Statistical Association, 100(469):109–120, 2005.*

Y. Chen, I.H. Dinwoodie, and S. Sullivant. Sequential Importance Sampling for Multiway Tables. The Annals of Statistics, 34(1):523–545, 2006.

Y. Chen and D. Small. Exact Tests for the Rasch Model via Sequential Importance Sampling. *Psychometrika, 70(1):11–30, 2005.*

G.N. Masters. A Rasch Model for Partial Credit Scoring. Psychometrika, 47(2):149–174, 1982.

I. Ponocny. Nonparametric goodness-of fit tests for the Rasch model. *Psychometrika, 66:437–460, 2001.*

T. Snijders. Enumeration and Simulation Methods for 0-1 Matrices with Given Marginals. Psychometrika, 56(3):397–417, 1991.

N.D. Verhelst. An efficient MCMC algorithm to sample binary matrices with fixed marginals. *Psychometrika, 73:705–728, 2008.*