A new procedure for detecting departures from local independence in item response models

Michael C. Edwards[1] Li Cai[2]

[1]Department of Psychology
The Ohio State University

[2]Graduate School of Education and Information Studies
UCLA

August 4, 2011

Today's Topics

- Local dependence
- An observation
- A new LD diagnostic
- Some demonstrations

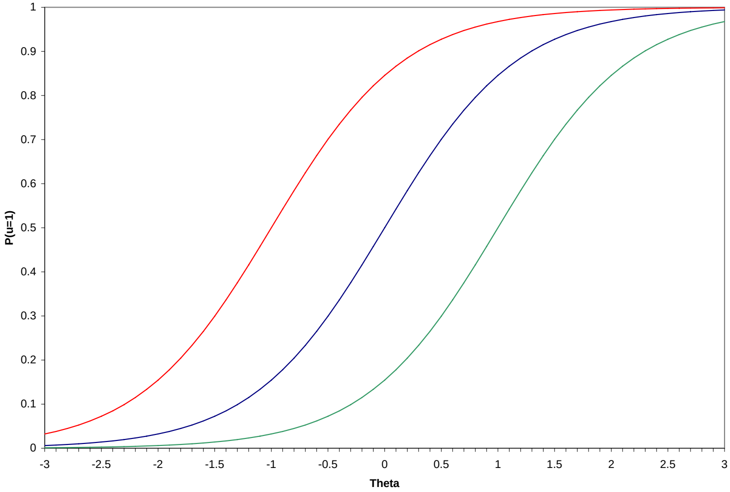
IRT in three slides - 1 of 3

One of the most widely used IRT models, the 2PL model is appropriate for dichotomous responses.

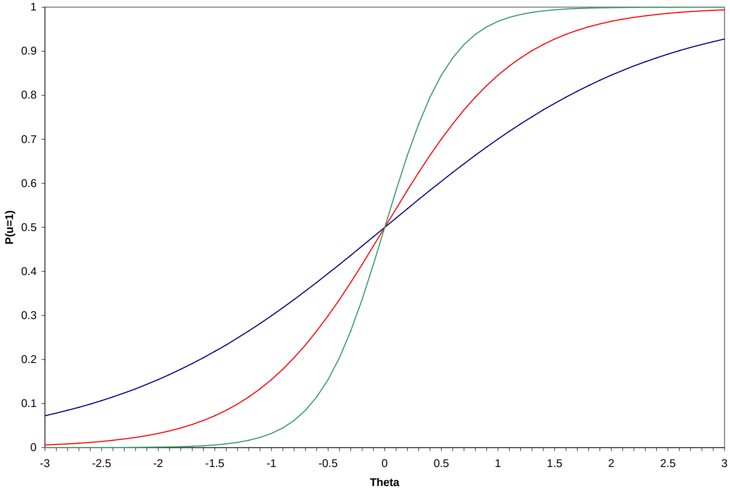
$$P(u_j = 1|\theta) = \frac{1}{1 + \text{exp}[-a_j(\theta - b_j)]} \tag{1}$$

- u_j is the observed response to item j
- a_j is the slope for item j
- b_j is the location parameter for item j
- θ is the latent construct being measured

IRT in three slides - 2 of 3



IRT in three slides - 3 of 3



IRT and dimensionality

Most IRT models we see are unidimensional. To use a unidimensional model we must assume that one and only one common factor gave rise to the observed item responses. If this is true, the resulting item responses are said to be locally independent - uncorrelated to one another after conditioning on the latent factor.

IRT and dimensionality continued

The truth is that this isn't really a unidimensional vs. not unidimensional issue. It's more of a "correctly specified number of common factors" vs. "incorrectly specified number of common factors".

IRT and dimensionality continued

The truth is that this isn't really a unidimensional vs. not unidimensional issue. It's more of a "correctly specified number of common factors" vs. "incorrectly specified number of common factors".

If there were two common factors, but they were modeled correctly, the item responses would still be conditionally (locally) independent.

Local Dependence

Simply put, local dependence (LD) is what you get when you fail to obtain local independence. It means that even after conditioning on the latent factor(s), some subset of the item responses are correlated.

Local Dependence

Simply put, local dependence (LD) is what you get when you fail to obtain local independence. It means that even after conditioning on the latent factor(s), some subset of the item responses are correlated.

If you're a factor analyst, LD occurs when your solution is under-factored. There are common factors you have not accounted for in your data.

Why is LD bad?

- Presence of LD means you're violating a major assumption
- LD can cause validity issues by "hijacking theta"
- Items exhibiting LD often have inflated slopes

An observation

In many cases where a pair of items are exhibiting LD, when they are both in the model the slopes for those two items are inflated. When one of the pair is removed (or the LD is dealt with in some other manner) the slope on the remaining item usually drops quite a bit.

An observation

In many cases where a pair of items are exhibiting LD, when they are both in the model the slopes for those two items are inflated. When one of the pair is removed (or the LD is dealt with in some other manner) the slope on the remaining item usually drops quite a bit.

This got me thinking...

An idea begins to form...

If a set of items is unidimensional, removing any individual item should have virtually no impact on the slopes of the remaining items.

On the other hand, if the item removed is locally dependent, then we might expect to see a fairly significant change in the slope of the remaining offender.

That sounds familiar

This is a lot like DFBETAS in OLS regression. If an observation is influential, then removing it will jerk the estimated slope and intercept around. If it isn't, then removing it shouldn't do much to the estimates.

The basic idea

We calculate all the slope parameters and then, one at a time, omit an item and re-analyze the remaining items.

The basic idea

We calculate all the slope parameters and then, one at a time, omit an item and re-analyze the remaining items.

For each item we take a difference between the “full set” slope and the “minus one” slope and divide it by the standard error of the “minus one” slope.

Calculating the diagnostic

The diagnostic is computed as

$$\frac{a_j - a_{j(k)}}{se(a_{j(k)})},$$

where a is the IRT slope estimate, j indexes item impacted and k indexes item removed. Each item receives a vector of K-1 diagnostics, one calculated with the removal of each other item.

We've been folding the matrix over the diagonal and adding the corresponding elements. This makes wayward pairs easier to spot and helps suppress some of the noise.

Some simulated examples

To get a sense for how this diagnostic performs we simulated three different conditions:

- Surface LD (SLD)
  - $\pi = 0.3$
- Underlying LD (ULD)
  - 2^{nd} factor with slopes 1.5 times larger than general slopes (2.55 and 1.7, respectively)

SLD original LD matrix (N=1,000, items 2 & 3)

Item Impacted	Item Removed 1	2	3	4	5	6	7	8	9	10
1		-0.2	-0.2	-0.4	0.2	0.2	0.1	0.0	0.0	0.4
2	-0.2		2.3	-0.2	-0.7	-0.7	-0.6	-0.7	-0.1	-0.4
3	-0.3	2.0		-0.1	-1.1	-0.1	-0.1	-1.0	-0.2	-0.3
4	-0.5	-0.3	-0.1		0.8	-0.2	-0.1	0.7	0.0	-0.3
5	0.2	-0.5	-0.9	0.6		-0.1	0.0	0.7	0.0	-0.1
6	0.3	-0.7	-0.2	-0.2	-0.1		0.5	0.1	0.1	0.4
7	0.1	-0.4	-0.2	-0.1	0.1	0.4		0.1	-0.1	0.4
8	-0.1	-0.5	-0.7	0.5	0.7	0.0	0.0		0.1	-0.3
9	0.1	-0.1	-0.1	0.0	0.1	0.1	-0.1	0.1		0.2
10	0.4	-0.4	-0.1	-0.2	-0.1	0.3	0.4	-0.2	0.1	

SLD folded LD matrix (N=1,000, items 2 & 3)

	Item									
Item	1	2	3	4	5	6	7	8	9	10
1		-0.4	-0.5	-0.9	0.4	0.5	0.2	-0.1	0.1	0.8
2			4.3	-0.5	-1.2	-1.4	-1.1	-1.2	-0.1	-0.8
3				-0.3	-1.9	-0.3	-0.3	-1.7	-0.3	-0.4
4					1.4	-0.4	-0.2	1.3	-0.1	-0.5
5						-0.2	0.1	1.5	0.1	-0.1
6							0.9	0.1	0.2	0.6
7								0.2	-0.2	0.8
8									0.2	-0.5
9										0.3

ULD folded LD matrix (N=1,000, items 1 & 2)

	Item									
Item	1	2	3	4	5	6	7	8	9	10
1		2.7	-0.6	-0.9	-0.4	-0.7	-0.3	0.0	-0.1	-0.1
2			-1.8	-0.8	-0.9	-0.5	0.1	0.2	1.0	-0.8
3				0.4	0.9	0.5	0.1	-0.1	0.8	-0.3
4					0.9	0.3	0.1	0.2	-0.5	0.3
5						0.2	-0.6	0.6	-1.5	0.8
6							-0.1	-0.8	0.0	0.4
7								0.0	0.8	0.3
8									-0.2	-0.1
9										-0.3

Preliminary conclusions

From the simulations conducted so far we've learned a few things:

- The diagnostic works better with larger N (smaller standard errors)
- As Chen & Thissen (1997) found, SLD is easier to detect than ULD (and we have some ideas why)
- We were able to find SLD and ULD with 1,000 subjects, even with relatively weak LD

A slightly more complex real data example

The simulated examples only involved pairs of items which were locally dependent (a.k.a. doublets). Since any kind of unmodeled multidimensionality represents a departure from local independence, we were curious to see if this LD diagnostic could be useful in more complex situations.

A slightly more complex real data example

This example comes from a Quality of Life scale developed by Lehman (1988) and analyzed to illustrate full-information bi-factor analysis of graded response data by Gibbons et al. in a 2007 *Applied Psychological Measurement* paper.

A slightly more complex real data example

The Quality of Life Rating Scale (QoLRS) consists of 35 items which are divided fairly evenly among seven subscales: Family, Finance, Health, Leisure, Living, Safety, and Social.

Each item has a 7-point response alternative. The data for this example (very kindly provided by Dr. Lehman) has 586 subjects' responses to all 35 items.

Bi-Factor Loadings for QoLRS

	1	2	3	4	5	6
1	0.7	0				
2	0.5		0.6			
3	0.5		0.4			
4	0.6		0.5			
5	0.6		0.5			
6	0.4			0.6		
7	0.4			0.5		
8	0.5			0.6		
9	0.5			0.6		
10	0.5				0.3	
11	0.5				0.4	
12	0.4				0.4	
13	0.5				0.4	
14	0.6				0.3	
15	0.6				0.2	
16	0.6					0.3
17	0.5					0.3
18	0.6					0.4
19	0.6					0.4
20	0.5					0.4
21	0.4					0.2

	1	7	8	9
22	0.5	0.5		
23	0.4	0.5		
24	0.5	0.6		
25	0.5	0.6		
26	0.5	0.6		
27	0.5		0.4	
28	0.5		0.4	
29	0.5		0.4	
30	0.5		0.5	
31	0.5		0.3	
32	0.5			0.3
33	0.5			0.3
34	0.5			0.3
35	0.4			0.2

LD diagnostic and the QoLRS

The resulting 35 x 35 matrix of diagnostics is a bit cumbersome (1190 elements).

The matrix is presented (albeit a bit small) on the next slide, with the largest 10% highlighted.

LD diagnostic and the QoLRS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1		0.1	0.0	0.0	0.2	-0.1	0.0	0.0	-0.1	0.1	0.0	-0.1	-0.1	0.2	0.4	0.2	-0.1	0.1	0.0	0.0	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	-0.1	0.0
2	0.1		0.4	0.6	0.6	-0.1	-0.1	-0.2	-0.2	-0.1	0.0	0.0	0.0	0.0	0.0	-0.2	-0.1	-0.1	-0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	-0.1	0.0	-0.1	0.0	0.0
3	0.0	0.4		0.5	0.6	0.0	0.0	-0.1	-0.1	-0.1	0.0	0.0	0.0	-0.1	-0.1	-0.1	0.0	-0.2	-0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0	0.1	0.0	0.0
4	0.0	0.5	0.4		0.7	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	0.0	-0.1	-0.1	-0.2	-0.1	-0.1	-0.2	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	0.0
5	0.1	0.5	0.5	0.7		-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.2	-0.1	-0.1	-0.2	0.0	-0.1	0.0	0.0	0.1	0.0	0.0	-0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.1	0.0	0.0
6	-0.1	-0.1	0.0	-0.1	-0.1		0.3	0.5	0.5	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	0.0	-0.1
7	-0.1	-0.1	0.0	-0.1	-0.1	0.4		0.4	0.4	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	-0.1	-0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.1	0.1	-0.1	-0.1
8	0.0	-0.2	-0.1	-0.1	-0.1	0.5	0.3		0.5	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.1	0.0	0.0	0.1	0.1	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.0	-0.1	0.0	-0.1	-0.1	-0.1	0.1	0.1	-0.1
9	-0.1	-0.2	0.0	-0.1	-0.1	0.5	0.3	0.5		-0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.0	-0.1	-0.1	0.0	-0.1	0.0	-0.1	-0.2	-0.1	-0.1	-0.1	0.0	0.1	0.0	-0.1
10	0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	-0.2	-0.1		0.1	0.1	0.0	0.5	0.2	-0.1	0.0	0.0	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	-0.1	-0.2	-0.1	0.0
11	0.0	0.0	0.0	-0.1	-0.2	0.0	0.0	0.0	0.0	0.1		0.3	0.3	0.2	0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.1	-0.1	0.0	0.0	0.0	-0.1	-0.1	0.0	0.1	0.0	0.0	0.0	-0.1	0.0	0.0
12	-0.2	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.1	0.3		0.3	0.1	0.1	-0.1	0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	0.0	-0.1	0.0	-0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.0
13	-0.1	0.0	0.0	0.0	-0.1	-0.1	0.0	-0.1	-0.1	0.0	0.3	0.3		0.1	0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.0	-0.1	0.0	0.0
14	0.2	0.0	0.0	-0.1	-0.1	0.0	0.0	-0.1	-0.1	0.4	0.2	0.1	0.1		0.2	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	-0.1	0.0	0.0	0.1	0.0	-0.1	-0.1	0.0	0.0
15	0.5	0.0	-0.1	-0.1	0.0	-0.1	0.0	-0.1	0.0	0.1	0.1	0.1	0.1	0.3		0.1	0.0	0.0	0.0	0.0	-0.1	-0.2	-0.1	-0.2	-0.2	-0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
16	0.2	-0.1	-0.1	-0.2	-0.2	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	-0.1	0.0	0.1		0.3	0.2	0.4	0.2	0.0	-0.2	-0.1	-0.2	-0.2	-0.1	0.1	-0.1	0.0	-0.1	0.1	0.2	0.2	0.2	0.0
17	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.4		0.3	0.3	0.3	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.1	0.1	-0.1
18	0.1	-0.1	-0.2	-0.2	-0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.3	0.3		0.5	0.2	0.0	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.2	0.0	-0.1	0.0	0.0	0.1	0.0	0.0
19	0.0	-0.2	-0.2	-0.2	-0.2	0.0	0.0	0.1	0.1	-0.1	-0.1	0.0	-0.1	-0.1	0.0	0.4	0.2	0.4		0.4	0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	0.0	-0.1	-0.1	0.1	0.2	0.1	0.0
20	-0.1	-0.1	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.2	0.3	0.2	0.4		0.1	-0.1	-0.1	-0.1	0.0	0.0	-0.1	-0.2	0.0	-0.1	-0.1	0.0	0.1	0.0	-0.1
21	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	0.0	-0.1	0.1	0.0	0.0	0.0	-0.1	-0.1	0.0	0.1	0.1	0.1		0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.2	-0.2	-0.1	-0.1	-0.1	-0.1	0.0		0.3	0.3	0.3	0.3	-0.1	0.2	-0.1	0.0	0.0	0.0	-0.1	-0.1	0.0
23	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	0.0	0.0	-0.1	0.0	-0.1	-0.2	-0.1	-0.1	0.0	-0.2	-0.1	0.1	0.3		0.2	0.3	0.2	0.0	0.1	-0.1	0.0	0.1	0.0	-0.1	-0.1	0.0
24	-0.1	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	-0.2	-0.2	-0.1	-0.1	-0.1	-0.1	0.1	0.4	0.2		0.4	0.3	-0.2	0.1	-0.1	0.0	0.0	-0.1	-0.1	-0.1	0.0
25	-0.1	0.0	0.0	0.0	0.0	-0.1	0.0	-0.1	-0.1	0.0	0.0	0.0	0.1	0.0	-0.2	-0.2	0.0	-0.1	-0.1	0.0	0.0	0.3	0.2	0.4		0.3	-0.1	0.1	-0.1	0.0	0.1	0.0	-0.1	-0.1	0.1
26	-0.1	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.1	0.0	-0.1	0.0	0.0	-0.1	-0.1	-0.1	-0.1	-0.2	0.0	0.0	0.4	0.2	0.3	0.3		-0.1	0.1	-0.1	0.0	0.0	-0.1	-0.1	-0.1	0.0
27	0.0	-0.1	-0.1	0.0	-0.1	-0.1	0.0	0.0	-0.1	0.1	-0.1	0.0	0.0	0.0	0.0	0.1	-0.1	0.0	-0.1	-0.1	0.0	0.0	0.0	-0.1	-0.1	-0.1		0.4	0.2	0.4	0.4	-0.1	-0.1	0.0	0.0
28	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	-0.2	0.0	0.0	-0.1	0.0	0.0	0.0	-0.1	-0.1	-0.2	-0.2	-0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.4		0.1	0.4	0.4	-0.1	-0.2	-0.1	0.1
29	0.0	0.0	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	-0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.2	-0.1	-0.1	0.3	0.2		0.3	0.2	0.0	-0.1	0.0	0.1
30	-0.1	-0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	0.1	0.0	0.1	0.0	0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.3		0.3	-0.1	-0.2	-0.1	0.0
31	0.0	-0.1	-0.2	-0.1	-0.2	-0.1	-0.1	-0.1	-0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.0	-0.1	-0.2	0.0	0.0	0.1	0.0	0.1	0.0	0.4	0.4	0.1	0.3		0.0	-0.2	0.0	0.1
32	0.0	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.1	0.0	-0.1	0.0	0.2	0.0	0.0	0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	0.0		0.5	0.3	0.1
33	0.0	-0.1	0.1	-0.1	-0.1	0.0	0.1	0.1	0.1	-0.2	-0.1	0.0	-0.1	-0.1	0.0	0.2	0.1	0.1	0.2	0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	-0.1	-0.1	-0.2	0.4		0.3	0.0
34	-0.1	-0.1	0.0	-0.1	-0.1	0.0	-0.1	0.0	0.0	-0.1	0.0	0.1	0.0	0.0	0.0	0.2	0.1	-0.1	0.1	0.0	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.0	0.4	0.4		0.1
35	0.0	0.0	0.0	0.1	0.0	-0.2	-0.1	-0.2	-0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	-0.1	0.0	-0.1	-0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.2	0.1	0.2	

Not too shabby

We were pleasantly surprised by how well this LD diagnostic seemed to do with the dimensionally complex QoLRS.

More preliminary conclusions

To sum up, the LD diagnostic we propose here seems to do a good job identifying doublet-type LD and, based on the QoLRS analysis, may actually have some promise with respect to identifying more complex kinds of multidimensionality.

The End

Thank you.

edwards.134@osu.edu