A Primer on Design Matrices

Cathleen Kennedy

DRAFT of 3-5-03

Item Response Probability

To estimate a student's proficiency from his or her responses on an assessment, we use a formulation that evaluates the probability of a person with a known ability responding in each category to an item with known difficulties.

When an item response has only two possible values, correct or incorrect, the item difficulty is an expression of how much more ability a person needs to get a correct answer rather than an incorrect answer. More precisely, we describe the item difficulty as the ability level where the student is equally likely to get a correct or incorrect response (that is, both are .5). In Figure 1 the item difficulty is approximately –0.6.

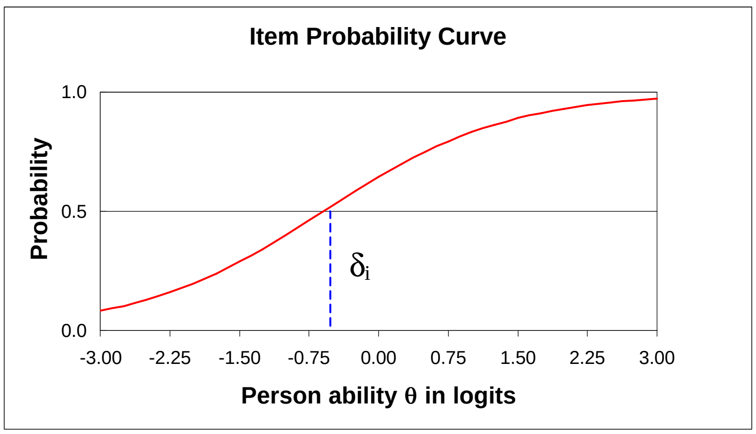


Figure 1 – Probability curve for a dichotomous item.

When items have more than two possible outcomes, we need more information than just the item difficulty. We need to know how much more ability is needed to achieve each possible score on the item. The first of these is parallel to the description above, the ability required to have a 50-50 chance of get the item correct at any level rather than getting the item completely wrong. Subsequent steps in difficulty are interpreted in much the same way. The second step difficulty is the ability required to have a 50-50 chance of getting a score of 2 rather than a score of 1 on the item.

Units of Measurement

The scale of measurement is defined in terms of *logits*, which is a transformation of probability, π. A logit is the log of the odds of a particular event compared to its complement. (Note that $\exp(\lambda)$ means e^{λ} in the equations that follow):

$$p = \frac{\exp(l)}{(1+\exp(l))}, \text{ and}$$

$$l = \log\left[\frac{p}{(1-p)}\right].$$

Mathematical Models

To compute the probability of earning a score of j rather than j-1 on item i, given a set of fixed item parameters for the item, ξ_i, and a specific student proficiency of θ (in the unidimensional case), we use a Rasch formulation in the form:

$$P(x_i = j \mid q, x_i) = \frac{P(x = j \mid q, x_i)}{P(x = j-1 \mid q, x_i) + P(x = j \mid q, x_i)} = \frac{\exp(q - x_i)}{1 + \exp(q - x_i)} \qquad (1)$$

For the dichotomous case there are only two categories, j=0 and j=1, and P(x=0) + P(x=1)=1. If we represent the item difficulty of the item as δ_1 (i.e. ξ_1 is a single value), from (1) we have the equations:

$$P(x = 0) = \frac{1}{1 + \exp(q - d_1)}$$

$$P(x = 1) = \frac{\exp(q - d_1)}{1 + \exp(q - d_1)}$$

The partial credit case is an extension of the dichotomous case. Moving from one category to another implies a dichotomous choice between the lower and upper levels. What we really want to know is the probability of getting a particular response rather than all the others, so we can modify the probability equation to represent that value. The following equation shows the probability that a person with ability θ will respond in category *s* rather than in any other category on item i, given item difficulty parameters $\xi_i = \delta_{i1}, \delta_{i2}, \ldots \delta_{im}$.

$$P(x_i = s \mid q, x_i) = \frac{\exp \sum_{j=0}^{s} (q - d_{ij})}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} (q - d_{ij})}, \text{ where } s=0,1,2,\ldots,m \text{ (i.e. count of completed steps)}$$

Thus, for a 3-category item,

$$P(x = 0) = \frac{1}{\sum_{k=0}^{2} \exp \sum_{j=0}^{k} (q - d_{ij})}$$

$$P(x = 1) = \frac{\exp \sum_{j=0}^{1} (q - d_{ij})}{\sum_{k=0}^{2} \exp \sum_{j=0}^{k} (q - d_{ij})} = \frac{\exp(q - d_{i1})}{\sum_{k=0}^{2} \exp \sum_{j=0}^{k} (q - d_{ij})}$$

$$P(x = 2) = \frac{\exp \sum_{j=0}^{2} (q - d_{ij})}{\sum_{k=0}^{2} \exp \sum_{j=0}^{k} (q - d_{ij})} = \frac{\exp(q - d_{i1} + q - d_{i2})}{\sum_{k=0}^{2} \exp \sum_{j=0}^{k} (q - d_{ij})} = \frac{\exp(2q - (d_{i1} + d_{i2}))}{\sum_{k=0}^{2} \exp \sum_{j=0}^{k} (q - d_{ij})}$$

Note the convention $\exp(0) \equiv 1$ and $\sum_{j=0}^{0} q - d_{ij}) \equiv 0$;

$\sum_{k=0}^{m} \exp \sum_{j=0}^{k} (q - d_{ij})$ is the sum of the numerators for all categories.

Note the patterns in the numerators for the series of equations for each response category of a single item. For example, in the partial credit case with a 3-category item:

P(x=0): $\exp(0\theta - 0)$

P(x=1): $\exp(1\theta - 1\delta_{i1})$

P(x=2): $\exp(2\theta - (1\delta_{i1} + 1\delta_{i2}))$

The θ coefficients comprise the entries in a Scoring Matrix, B:

$$B = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

The δ coefficients comprise the entries in a Design Matrix, A:

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix},$$

where column 1 represents the δ_{i1} coefficients and column 2 represents the δ_{i2} coefficients.

Figure 2 – Correlations of coefficients with Scoring and Design matrices.

In the partial credit model, we parameterize the difficulty of achieving a score of j on item i and represent it with δ_{ij}. That is, δ_{ij} is the ability level one would require to expect a 50-50 chance of responding in category j rather than in category j-1. We might think of the average of the δ_{ij}'s as an overall item difficulty, and the step difficulties as each step's deviance from the average. In looking at item difficulties in this way we are saying that each δ_{ij} is a composite of $\delta_i + \tau_{ij}$, where τ_{ij} is the deviance from the average item difficulty for item *i* at step *j*.

The rating scale model is a special case of the partial credit model in which the τ values are the same for every item. That is, $\tau_{11}=\tau_{21}=\tau_{31}$... and $\tau_{12}=\tau_{22}=\tau_{32}$.... In this model, $\xi_i = \delta_i, \tau_1, \ldots \tau_m$. In this formulation, our mathematical model becomes

$$P(x_i = s \mid q, x_i) = \frac{\exp \sum_{j=0}^{s} [q - (d_i + t_j)]}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} [q - (d_i + t_j)]}$$

Multidimensional Random Coefficients Multinomial Logit (MRCML) Model

The MRCML family of measurement models uses two matrices to facilitate the computations: 1) a Scoring Matrix that represents the relationships between items and their categories (the rows) and dimensions, or student model variables, (the columns); and 2) a Design Matrix that represents the relationships between items and categories (the rows) and the model parameters (e.g. item difficulties, step difficulties, etc.).

The Scoring Matrix is used to construct the θ component of the equations while the Design Matrix is used to construct the δ part of the equations.

Examples

The following are a series of models demonstrating how one might construct the matrices in their cases:

1. Unidimensional Dichotomous model (e.g. true/false, multiple choice item).

Item Scoring Matrix (one dimension, so one column):

	D_1
category 1	0
category 2	1

Item Design Matrix (one item difficulty needed, so one column):

	δ_i
category 1	0
category 2	1

In the case of an assessment with 10 dichotomous items, the associated *assessment* scoring and design matrices would be:

Assessment Scoring Matrix:

	D_1
item 1, category 1	0
item 1, category 2	1
item 2, category 1	0
item 2, category 2	1
:	:
:	:
item 10, category 1	0
item 10, category 2	1

Assessment Design Matrix:

	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8	δ_9	δ_{10}
item 1, category 1	0	0	0	0	0	0	0	0	0	0
item 1, category 2	1	0	0	0	0	0	0	0	0	0
item 2, category 1	0	0	0	0	0	0	0	0	0	0
item 2, category 2	0	1	0	0	0	0	0	0	0	0
:	:									
:	:									
item 10, category 1	0	0	0	0	0	0	0	0	0	0
item 10, category 2	0	0	0	0	0	0	0	0	0	1

2. Unidimensional Partial Credit model (e.g. different responses earn different points).
For a single item with 4 categories, we would have:

Item Scoring Matrix (one dimension, so one column):

	D_1
category 1	0
category 2	1
category 3	2
category 4	3

We might describe a particular score as, "A response in category 3 has a score of 2."

Item Design Matrix (four categories means three steps, so 3 columns):

	δ_{i1}	δ_{i2}	δ_{i3}
category 1	0	0	0
category 2	1	0	0
category 3	1	1	0
category 4	1	1	1

Here we would say, "The difficulty of getting a response in category 3 is computed from the difficulty of advancing from category 1 to category 2 (the first column), and the difficulty in advancing from category 2 to category 3 (the second column)." The difficulty of achieving a response in category 3 is conditional on being able to earn the lower scores also. That is, getting a score of 3 means the student earned a 3 given he or she had already earned a 2.

For an assessment with 5 items in which items 1 through 3 are 5-category items and items 4 and 5 have 3 categories, the matrices would be:

Assessment Scoring Matrix:

	D_1
item 1, category 1	0
item 1, category 2	1
item 1, category 3	2
item 1, category 4	3
item 1, category 5	4
:	:
:	:
item 5, category 1	0
item 5, category 2	1
item 5, category 3	2

Assessment Design Matrix:

	δ_{11}	δ_{12}	δ_{13}	δ_{14}	δ_{21}	δ_{22}	δ_{23}	δ_{24}	δ_{31}	δ_{32}	δ_{33}	δ_{34}	δ_{41}	δ_{42}	δ_{51}	δ_{52}
item 1, category 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
item 1, category 2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
item 1, category 3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
item 1, category 4	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
item 1, category 5	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
:	:															
:	:															
item 5, category 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
item 5, category 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
item 5, category 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

3. Unidimensional Rating Scale model (e.g. a Likert-type item) with 5 categories.
For a single item with 4 categories the mathematical model is the same as for the partial credit case; however, since we parameterize the item difficulty differently in this model, we construct the design matrix differently also (the need for this is more apparent for an entire assessment than for a single item).

Item Scoring Matrix (one dimension, so one column):

	D_1
category 1	0
category 2	1
category 3	2
category 4	3
category 5	4

We might describe a particular score as, "A response in category 3 has a score of 2."

Item Design Matrix (item difficulty and four step difficulties, so 5 columns):

	δ_i	τ_1	τ_2	τ_3	τ_4
category 1	0	0	0	0	0
category 2	1	1	0	0	0
category 3	2	1	1	0	0
category 4	3	1	1	1	0
category 5	4	1	1	1	1

Here we would say, "The difficulty of getting a response in category 3 is computed from the difficulty of the item (the first column), the deviance from the item difficulty to get a response in category 2 (the second column), and the deviance from the item difficulty to get a response in category 3 (the third column)." These step parameters have a different interpretation, and are computed differently, from the step parameters in the partial credit model, so the formulation of the design matrix looks different from the partial credit model.

Note that the total difficulty of getting a category 3 response is:
(item diff. + diff. in going from cat. 1 to 2) + (item diff. + diff in going from cat. 2 to 3),
or 2*(item diff.) + cat.1 to 2 + cat. 2 to 3

Since the sum of all the "step" difficulties (τs) is 0, the total difficulty of getting a category 5 response is 4*(item diff.), because cat. 1 to 2 + cat. 2 to 3 + cat. 3 to 4 + cat. 4 to 5 = 0. Thus, we can set the final row of τ parameters to 0 and eliminate the last parameter column.

Item Design Matrix:

	δ_i	τ_1	τ_2	τ_3
category 1	0	0	0	0
category 2	1	1	0	0
category 3	2	1	1	0
category 4	3	1	1	1
category 5	4	0	0	0

The rating scale model is considered a special case of the partial credit model. All items noted as Rating Scale in the PADI system must use the same parameter estimates for the steps. Some part of the system logic will have to verify this consistency.

For an assessment comprised of 5 rating scale items with 3 categories each, we would need:

Assessment Scoring Matrix:

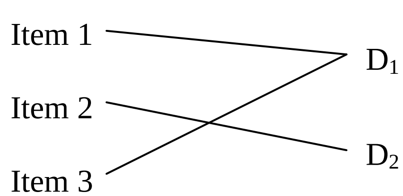
	D_1
item 1, category 1	0
item 1, category 2	1
item 1, category 3	2
item 2, category 1	0
item 2, category 2	1
:	:
:	:
item 5, category 1	0
item 5, category 2	1
item 5, category 3	2

Assessment Design Matrix:

	δ_1	δ_2	δ_3	δ_4	δ_5	τ_1	τ_2
item 1, category 1	0	0	0	0	0	0	0
item 1, category 2	1	0	0	0	0	1	0
item 1, category 3	2	0	0	0	0	1	1
item 2, category 1	0	0	0	0	0	0	0
item 2, category 2	1	0	0	0	0	1	0
:	:						
:	:						
item 5, category 1	0	0	0	0	0	0	0
item 5, category 2	0	0	0	0	1	1	0
item 5, category 3	0	0	0	0	2	1	1

4. **Between-item Multidimensional Partial Credit model (e.g. different responses earn different points) with 4 categories.**

Here, each item only maps to a single dimension, so representing a single item would be the same as the unidimensional case (example 2.). However, when we construct an assessment that measures proficiency on two or more Student Model Variables, we can need to relate individual items to specific Ds. In the following example, we have an assessment with 3 items. Items 1 and 2 have 3 categories, while item 3 has 4 categories. Items 1 and 3 map to D_1 and item 2 maps to D_2



Assessment Scoring Matrix:

	D_1	D_2
item 1, category 1	0	0
item 1, category 2	1	0
item 1, category 3	2	0
item 2, category 1	0	0
item 2, category 2	0	1
item 2, category 3	0	2
item 3, category 1	0	0
item 3, category 2	1	0
item 3, category 3	2	0
item 3, category 4	3	0

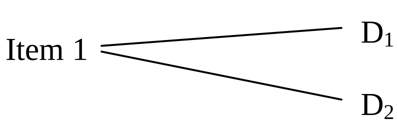
Assessment Design Matrix:

	δ_{11}	δ_{12}	δ_{21}	δ_{22}	δ_{31}	δ_{32}	δ_{33}
item 1, category 1	0	0	0	0	0	0	0
item 1, category 2	1	0	0	0	0	0	0
item 1, category 3	1	1	0	0	0	0	0
item 2, category 1	0	0	0	0	0	0	0
item 2, category 2	0	0	1	0	0	0	0
item 2, category 3	0	0	1	1	0	0	0
item 3, category 1	0	0	0	0	0	0	0
item 3, category 2	0	0	0	0	1	0	0
item 3, category 3	0	0	0	0	1	1	0
item 3, category 4	0	0	0	0	1	1	1

5. **Within-item Multidimensional Partial Credit model.**

Let's take the example where the item maps to two dimensions, or student model variables. One way to score this is to give two scores, one for the first dimension and one for the second. In this case, there will be more categories because we are considering combinations of scores on the two dimensions. In this example, the first dimension, D_1 has 4 categories while the second dimension, D_2 has only 3 categories.

The overall item category 1 means the student had a response in category 1 on the first dimension and a response in category 1 on the second dimension. We construct the complete set of item categories by building permutations on the combinations of scores on the two dimensions.



Item Scoring Matrix:

	D_1	D_2
category 1 (1,1)	0	0
category 2 (1,2)	0	1
category 3 (1,3)	0	2
category 4 (2,1)	1	0
category 5 (2,2)	1	1
category 6 (2,3)	1	2
:	:	:
category 12 (4,3)	3	2

Item Design Matrix:

	$\delta_{D1,1}$	$\delta_{D1,2}$	$\delta_{D1,3}$	$\delta_{D2,1}$	$\delta_{D2,2}$ ($\delta_{dimension,step}$)
category 1 (1,1)	0	0	0	0	0
category 2 (1,2)	0	0	0	1	0
category 3 (1,3)	0	0	0	1	1
category 4 (2,1)	1	0	0	0	0
category 5 (2,2)	1	0	0	1	0
category 6 (2,3)	1	0	0	1	1
:					
category 12 (4,3)	1	1	1	1	1

The design matrix may also need to change to reflect a more complex conceptualization of item difficulties that includes interaction effects between the two dimensions. For example, the first item parameter may represent the difficulty of the first dimension given a response in the first category on the second dimension. The second parameter may represent the difficulty of getting a response in the second category on the second dimension given a response in the first category on the first dimension.

Item Design Matrix:

	$\delta_{D1,1}\|x_{D2}=0$	$\delta_{D2,2}\|x_{D1}=0$	etc.
category 1 (1,1)	0	0	....
category 2 (1,2)	0	1	....
category 3 (1,3)	0	1	....
category 4 (2,1)	1	0	....
category 5 (2,2)	1	1	....
category 6 (2,3)	1	1	....
:	:	:	....
category 12 (4,3)	3	1	....

6. A Simple Item Bundle example:

3 dichotomous items in the bundle, each item maps to the same dimension.

The saturated model has 8 response patterns (the number of representations of 3 items with 2 categories each) to represent with a bundle-score:

```
category 1 0,0,0 = 0
category 2 1,0,0 = 1
category 3 0,1,0 = 2
category 4 0,0,1 = 3
category 5 1,1,0 = 4
category 6 0,1,1 = 5
category 7 1,0,1 = 6
category 8 1,1,1 = 7
```

The Item Bundle, rather than individual items, maps to the scoring matrix and the design matrix.

Item Bundle Scoring Matrix:

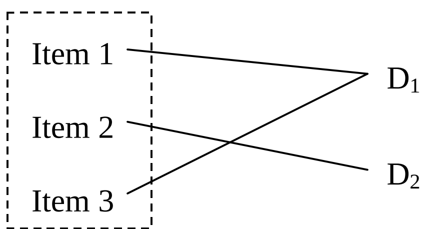
	D_1
category 1	0
category 2	1
category 3	1
category 4	1
category 5	2
category 6	2
category 7	2
category 8	3

Item Bundle Design Matrix of 8 rows (8 response patterns, or categories) and 7 columns (categories-1):

	δ_{i1}	δ_{i2}	δ_{i3}	δ_{i4}	δ_{i5}	δ_{i6}	δ_{i7}
category 1	0	0	0	0	0	0	0
category 2	1	0	0	0	0	0	0
:	0	1	0	0	0	0	0
	.	.	.	.	.	.	.
category 8	0	0	0	0	0	0	1

7. A "between-items" multidimensional item bundle example: the item bundle is multidimensional but each component item maps to only one dimension.

Now, let's assume that items 1 and 3 map to dimension D_1 and item 2 maps to dimension D_2. (Note that this model is similar to example 4, but here the items are conditionally dependent.)

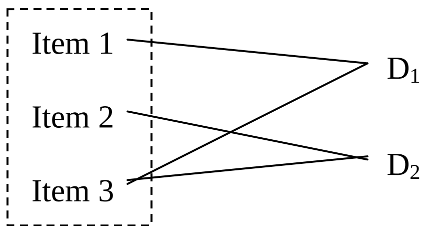


Then the Item Bundle Scoring Matrix would change to:

		D_1	D_2
category 1	0,0,0 = 0	0	0
category 2	1,0,0 = 1	1	0
category 3	0,1,0 = 2	0	1
category 4	0,0,1 = 3	1	0
category 5	1,1,0 = 4	1	1
category 6	0,1,1 = 5	1	1
category 7	1,0,1 = 6	2	0
category 8	1,1,1 = 7	2	1

8. Finally, a "within-item" multidimensional item bundle example: the item bundle is multidimensional and individual component items may map to multiple dimensions.

Now, let's assume that item 1 maps to dimension D_1, item 2 to dimension D_2, and item 3 to dimensions D_1 and D_2. (Note that this model is similar to example 5, but here the items are conditionally dependent.)



Item Bundle Scoring Matrix:

		D_1	D_2	
category 1	0,0,0 = 0	0	0	
category 2	1,0,0 = 1	1	0	
category 3	0,1,0 = 2	0	1	
category 4	0,0,1 = 3	1	1	
category 5	1,1,0 = 4	1	1	
category 6	0,1,1 = 5	1	2	(item 2 maps to D_2, item 3 maps to D_1 and D_2)
category 7	1,0,1 = 6	2	1	(item 1 maps to D_1, item 3 maps to D_1 and D_2)
category 8	1,1,1 = 7	2	2	(item 1 maps to D_1, item 2 maps to D_2, item 3 maps to D_1 and D_2)

The design matrix would also change to capture any interaction effects between the two dimensions (as in example 5 above)…. *this is left as an exercise for the reader*…

References:

Wilson, M. and Adams, R. J. (1995) Rasch Models for Item Bundles, *Psychometrika, 60 (2),* 181-198.

Wang, W.C., Wilson, M., Cheng, Y.Y. (2000?) *Local Dependence between Latent Traits when Common Stimuli are Used*, National Chung Cheng University.

Draney, K., Yamada, H., Xie, Y. (2000) *Cross-dimensional item bundling*, University of California at Berkeley.