Score Estimation and Generalised Partial Credit Models

Ray Adams and Greg Macaskill 30 September 2012

ConQuest 3 can estimate scoring parameters for a wide range of models. In the case of a unidimensional model for dichotomous data this model is well known as the two-parameter logistic model (Birnbaum, 1968). ConQuest 3 however goes well beyond this to include multidimensional forms of the two-parameter family of models, including multidimensional generalised partial credit models (Muraki , 1992) and multi-faceted models with score parameters estimated for each facet combination.

In ConQuest 3 we prefer to use the term *scores* to describe the additional parameters that are freed to be estimated.

Model Specification

As is fully described in Adams and Wu (2007) the ConQuest model is specified in two parts. The first part is a conditional categorical item response model and the second part is a population model. The item response model is commonly referred to as the mixed coefficients multinomial logit model (MCML).

Under the model the regression of the response vector on the item and person parameters is

$$f(\mathbf{x};\boldsymbol{\xi}|\boldsymbol{\theta}) \quad = \quad \Psi(\boldsymbol{\theta},\boldsymbol{\xi})\exp\left[\mathbf{x}^T(\mathbf{B}\boldsymbol{\theta}+\mathbf{A}\boldsymbol{\xi})\right], \tag{1}$$

with

$$\Psi(\boldsymbol{\theta},\boldsymbol{\xi})=\left\{\sum_{\mathbf{z}\in\Omega}\exp\left[\mathbf{z}^T(\mathbf{B}\boldsymbol{\theta}+\mathbf{A}\boldsymbol{\xi})\right]\right\}^{-1}, \tag{2}$$

where Ω is the set of all possible response vectors.

The dependent variable x is a vector-valued variable that describes the response pattern to a set of items. The model is referred to as a mixed coefficients model because items are described by a fixed set of unknown parameters, ξ, while the student outcome levels (the latent variable), $\boldsymbol{\theta}$, is a (multidimensional) random effect. The distributional assumptions for this random effect are specified through the population model.

Items are described through a vector, $\boldsymbol{\xi}^T=(\xi_1,\xi_2,\ldots,\xi_p)$, of *p* parameters. Linear combinations of these are used in the response probability model to describe the empirical characteristics of the response categories of each item. Design vectors, $\mathbf{a}_{ik},(i=1,\ldots,I;k=1,\ldots K_i)$, each of length *p,* which can be collected to form a design matrix $\mathbf{A}^T=(\mathbf{a}_{11},\mathbf{a}_{12},\ldots,\mathbf{a}_{1K_1},\mathbf{a}_{21},\ldots,\mathbf{a}_{2K_2},\ldots,\mathbf{a}_{IK_I})$, define these linear combinations.

The multi-dimensional form of the model assumes that a set of *D* traits underlies the individuals' responses. The *D* latent traits define a *D*-dimensional latent space. The vector $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_D)'$ represents an individual's position in the *D*-dimensional latent space.

The model also introduces a scoring function that allows the specification of the score or performance level assigned to each possible response category to each item. To do so, the notion of a response score b_{ikd} is introduced, which gives the performance level of an observed response in category *k*, item *i*, dimension *d*. The scores across *D* dimensions can be collected into a column vector $\mathbf{b}_{ik} = \left(b_{ik1}, b_{ik2}, \ldots, b_{ikD}\right)^T$ and again collected into the scoring sub-matrix for item *i*, $\mathbf{B}_i = \left(\mathbf{b}_{i1}, \mathbf{b}_{i2}, \ldots, \mathbf{b}_{iD}\right)^T$ and then into a scoring matrix $\mathbf{B} = \left(\mathbf{B}_1^T, \mathbf{B}_2^T, \ldots, \mathbf{B}_I^T\right)^T$ for the entire test.

If the scoring matrix is specified a-priori then (1) and (2) specify Rasch family models, whereas if the values in the scoring matrix are estimated from the data then the model is no longer a Rasch model. For simplicity we refer to a model with estimated scores as a 2PL model.

With appropriate choices for the design matrix **A** and with appropriate constraints on **B** the model given by (1) and (2) can be shown to be equivalent to many named item response models. For example, if the model is unidmensional, all items are dichotomous and **A** is an identity matrix (multiplied by –1) then (1) and (2) become, for a single item:

$$f(x_{ni}; b_i, \delta_i | \theta_n) = \frac{exp[x_{ni}(b_i\theta_n - \delta_i)]}{1 + exp(b_i\theta_n - \delta_i)}, \tag{3}$$

so that δ_i is the estimated item location parameter and b_i is the estimated score (or discrimination) parameter.

Similarly appropriate choices **A** and appropriate constraints on **B** can be chosen so that (1) and (2) become

$$f(x_{ni}; b_{i1}, b_{i2}, \ldots, b_{ik}, \delta_{i1}, \delta_{i2}, \ldots, \delta_{ik} | \theta_n) = \frac{exp\left(b_{ij}\theta_n - \sum_{t=1}^{j}\delta_{it}\right)}{\sum_{k=1}^{K} exp\left(b_{ik}\theta_n + \sum_{t=1}^{k}\delta_{it}\right)}, \tag{4}$$

which is a generalised partial credit model. Note, however that this form of the generalised partial, credit is somewhat more general than that proposed by Muraki (1992). In the Muraki model a single score, b_i, is estimated for the item and then the score for the *k-th* category is $\frac{k}{K} b_i$

Fitting the Models

The estimation of scores is accomplished by adding the `scoresfree` option to the `model` command. For identification purposes (see below) the use of case constraints (`set constraint=cases)` is also required.

If the `scoresfree` option is used then a score is estimated for every response category of every generalised item that is defined by the model. Recall that generalised items are defined by all the unique combinations of facets. For example, the `model item+rater` applied to dichotomous data would result in a score for each of the item and rater combinations.

As with all ConQuest 3 models the number of categories that are modelled is a function of the outcomes of scoring. The score values that are assigned to categories (via score, key and recode statements) are taken as initial values, with the exception of zero scores which are fixed at zero and are not free to be estimated.

Identification Requirements

For Rasch models (models with fixed **B** values) location constraints are required. In the case of single facet models this is typically achieved by fixing either the case mean to be zero or the item mean to be zero. If the **B** values are also estimated then a scale constraint is also required. In the case of unidimensional models the scale constraint is imposed by setting the variance of the latent variable to 1.0. In the case of multidimensional models variance-covariance matrix is constrained to have unit diagonals, that is, the estimated matrix is a correlation matrix.

In addition to scale and location constraints it is also necessary to have at least one category within each generalised item scored as zero. In the majority of applications this will be the (most) incorrect response category.

Limitation and Restrictions

At present the `scoresfree` option is not fully implemented for multi-dimensional models. The item parameters and their standard errors are estimated correctly as are the scores,, but the standard errors are incorrect. At present, standard errors in a between-item multidimensional model are given for those scores which are by definition zero, and in addition to this the `empirical` option gives infinite values.

Output of Results

If the estimated model includes scores then there are a number of implications for the program output. First, in the case of dichotomous items, an additional column is added to the item parameter estimates table. This additional column is equal to the δ_i parameter given in model (2) divided by the estimated score. This yields an outcome, in the case of dichotomous items, equivalent to the following parameterisation:

$$f(x_{ni}; b_i, \delta_i | \theta_n) = \frac{exp[x_{ni} b_i(\theta_n - \delta_i^*)]}{1 + exp[b_i(\theta_n - \delta_i^*)]}, \qquad (5)$$

Second, and additional output table is provided containing score parameter estimates for each category of each generalised item. The table can be obtained by the command: `show parameters!tables=11;`

Third, with the exception of predicted probability maps, item maps are not applicable for models with estimated scores.

Examples

A two parameter model can be easily fit to a multiple choice test as follows

```
set constraint=cases;
Format pid 1-5 responses 12-23;
Key acddbcebbacc ! 1;
Model item!scoresfree;
Estimate;
```

In this example two categories are modelled for each item. An incorrect response receives a score of zero for all items while the score for the correct response is estimated for each item.

A two parameter model can be easily fit to a multiple choice test as follows. In this example two categories are modelled for each item. An incorrect response receives a score of zero for all items while the score for the correct response is estimated for each item.

```
set constraint=cases;
```

```
Format pid 1-5 responses 12-23;
Key acddbcebbacc ! 1;
Model item!scoresfree;
Estimate;
```

An equivalent model can be achieved as follows. The code 'M' has been used in these data as a missing values indicator and in both the following and the previous syntax it will be scored as zero.

```
set constraint=cases;
Format pid 1-5 responses 12-23;
recode (a b c d e M) (1 0 0 0 0 0)!items(1,10);
recode (a b c d e M) (0 1 0 0 0 0)!items(5,8,9);
recode (a b c d e M) (0 0 1 0 0 0)!items(2,6,11,12);
recode (a b c d e M) (0 0 0 1 0 0)!items(3,4);
recode (a b c d e M) (0 0 0 0 1 0)!items(7);
Model item!scoresfree;
Estimate;
```

It is worthwhile contrasting the above two examples with the following.

```
set constraint=cases;
Format pid 1-5 responses 12-23;
score (a b c d e M) (1 0 0 0 0 0)!items(1,10);
score (a b c d e M) (0 1 0 0 0 0)!items(5,8,9);
score (a b c d e M) (0 0 1 0 0 0)!items(2,6,11,12);
score (a b c d e M) (0 0 0 1 0 0)!items(3,4);
score (a b c d e M) (0 0 0 0 1 0)!items(7);
Model item!scoresfree;
Estimate;
```

In this third example `score` rather than `recode` statements have been used. The absence of score statements in the first two examples means that a default scoring has been used. The default scoring simply maps the codes of '0' and '1' to scores of '0' and '1'. Two response categories are therefore modelled and the **B** values for the category coded '1' are estimated while the **B** for the categories scored '0' are fixed at zero. In the third example the score statement maps the codes in the data to six score categories and as a consequence all six categories are modelled with a 2PL version of the ordered partition model (Wilson and Adams, 1993).

An even more general model that might be considered is given in the following example

```
set constraint=cases;
Format pid 1-5 responses 12-23;
score (a b c d e M) (1 1 1 1 1 0)!items(1-12);
Model item!scoresfree;
Estimate;
```

In this fourth example scores are estimated for all categories of all items. One of the scores must be zero, and in this example the 'M' category is chosen. The scores of '1' that are assigned to all the other categories are treated as initial values.

As a fifth example we show a two-dimensional model with partial credit items and regression variables. Five categories are modelled for each item, but again note that the scores 1, 2, 3, and 4 are only initial values while the zero value is fixed.

```
format responses 1-9,50-114 grade 118 gender 119 ses 122-127!tasks(74);
model tasks+tasks*step!scoresfree;
recode (9) (0);
score (0,1,2,3,4) (0,1,2,3,4) ( ) !tasks(1-9);
score (0,1,2,3,4) ( ) (0,1,2,3,4) !tasks(10-74);
regression grade,gender,ses;
set constraint=cases;
```

```
estimate;
```

References

Adams, R. J., & Wu, M. L. (2007). The mixed-coefficient multinomial logit model: A generalized form of the Rasch model. In M. v. Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 57 – 76): Springer Verlag.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F.M. & Novick, M.R. (Eds.), Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Muraki, E. (1992). A generalized partial credit model. Applied Psychological Measurement, 16(2), 159-176.

Wilson, M.R., and Adams, R.J. (1993). Marginal maximum likelihood estimation for the ordered partition model. Journal of Educational Statistics. 18, 69-90.