Score Estimation and Generalised Partial Credit Models

Greg Macaskill and Ray Adams 14 June 2016

ConQuest 4 implements a more generalised item response model than that used in ConQuest 3 and described in note 6 (Adams & Macaskill, 2012). The ConQuest 3 model allowed the estimation of 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 also included 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 4 the model is further extended to allow constraints on the scoring parameters. The introduction of the constraints allows scoring parameters to be applied to multi-faceted models and it allows the estimation of latent variances that vary with group membership -- for example different latent variances for males and females or for two grade level.

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). Here we describe a more general form of the item response model that allows a more flexible approach to score parameter specification.

Under the ConQuest 3 model the regression of the response vector on the item and person parameters is:

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

with

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

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, $\xi^T=(\xi_1,\xi_2,\ldots,\xi_p)$, 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 matrix **B** 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 $\mathbf{b}_{ik} = (b_{ik1}, b_{ik2}, \ldots, b_{ikD})^T$ and again collected into the scoring sub-matrix for item i, $\mathbf{B}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, \ldots, \mathbf{b}_{iD})^T$ and then into a scoring matrix $\mathbf{B} = (\mathbf{B}_1^T, \mathbf{B}_2^T, \ldots, \mathbf{B}_I^T)^T$ for the entire test.

The ConQuest 4 model generalises the scoring by adding a vector $\boldsymbol{\tau}$ of s, scoring parameters, and a set of design matrices $\mathbf{C}_d$, one for each dimension. The dimension of these matrices is $K \times s$ where $K = \sum_{i=1}^{I}(K_i + 1)$ is the sum over the items of the number of response categories. This will allow estimation of scoring parameters and this form permits both different constraints for different dimensions and constraints across dimensions.

The probability of a response in category j of item i is then given as:

$$\Pr(X_{ij} = 1; A, B, \xi \mid \theta) = \frac{\exp(\sum_{d=1}^{D} \theta_d C_d \tau + a_{ij}^t \xi)}{\sum_{k=0}^{K_i} \exp(\sum_{d=1}^{D} \theta_d C_d \tau + a_{ik}^t \xi)} \quad (3)$$

and
$$f(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\tau} \mid \boldsymbol{\Theta}) = \Psi(\boldsymbol{\Theta}, \boldsymbol{\xi}, \boldsymbol{\tau}) \exp\left[\mathbf{x}^T \left(\sum_{d=1}^{D} \theta_d \mathbf{C}_d \boldsymbol{\tau} + \mathbf{A}\boldsymbol{\xi}\right)\right], \quad (4)$$

where
$$\Psi(\boldsymbol{\Theta}, \boldsymbol{\xi}, \boldsymbol{\tau}) = \left\{\sum_{\mathbf{z} \in \Omega} \exp\left[\mathbf{z}^T \left(\sum_{d=1}^{D} \theta_d \mathbf{C}_d \boldsymbol{\tau} + \mathbf{A}\boldsymbol{\xi}\right)\right]\right\}^{-1}$$

If the scoring parameters are specified a-priori then (3) and (4) 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.

With appropriate choices for the design matrix **A** and with appropriate choices for the **C** and $\boldsymbol{\tau}$ the model given by (3) and (4) can be shown to be equivalent to many named item response models. For example, if the model is unidmensional, all items are dichotomous, **A** is an identity matrix (multiplied by –1) and C is also an identity matrix then (3) and (4) become, for a single item:

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

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

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

$$f(x_{ni}; b_{i1}, b_{i2}, \dots, b_{ik}, \delta_{i1}, \delta_{i2}, \dots, \delta_{ik} | \theta_n) = \frac{exp\left(b_{ij}\theta_n - \sum_{t=1}^{j} \delta_{it}\right)}{\sum_{k=1}^{K} exp\left(b_{ik}\theta_n - \sum_{t=1}^{k} \delta_{it}\right)}, \qquad (6)$$

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 $\frac{k}{K} b_i$, whereas (6) allows parameters for each category.

As the item response model is conditional on the latent *D*-dimensional latent variable $\boldsymbol{\theta}$, we need to specify a distribution for $\boldsymbol{\theta}$ to complete the definition of the model. Using a multivariate normal distribution for the population we have

$$f_\Theta(\boldsymbol{\theta}; \mathbf{W}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) = (2\pi)^{-d/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\gamma}\mathbf{W})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\gamma}\mathbf{W})\right] \qquad (7)$$

where $\boldsymbol{\gamma}$ is a $U \times D$ matrix of regression coefficients, $\boldsymbol{\Sigma}$ is a $D \times D$ covariance matrix and $\mathbf{W}$ is a $U \times 1$ vector of fixed regression variables. With this population model, we have four sets of variables to be estimated, the population parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\gamma}$, the $P \times 1$ vector of item parameters $\boldsymbol{\xi}$ and the $S \times 1$ vector of score parameters $\boldsymbol{\tau}$.

The item response and population model together give the unconditional item response model

$$f(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) = \int_{\boldsymbol{\theta}} f_{\mathrm{x}}(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\tau} | \boldsymbol{\theta}) \, f_{\boldsymbol{\theta}}(\boldsymbol{\theta}; \boldsymbol{\gamma}, \boldsymbol{\Sigma}) \, d\boldsymbol{\theta}. \qquad (8)$$

Estimation

Maximum likelihood methods can be used to estimate the four sets of parameters in the model. The likelihood for *N* observations is then

$$\Lambda = \prod_{n=1}^{N} f(\mathbf{x}_n; \boldsymbol{\xi}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) \qquad (9)$$

If we define the marginal posterior as

$$h_\Theta(\boldsymbol{\theta}; \mathbf{W}, \boldsymbol{\tau}, \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}_g | \mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\tau}, \boldsymbol{\xi} | \boldsymbol{\theta}) \, f_\theta(\boldsymbol{\theta}; \mathbf{W}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})}{f_{\mathbf{x}}(\mathbf{x}; \mathbf{W}, \boldsymbol{\tau}, \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})}, \qquad (10)$$

After differentiation this leads to the following four sets of likelihood equations.

$$\sum_{n=1}^{N} \int_{\boldsymbol{\theta}} \left(\mathbf{x}_n^t - E(\mathbf{z}^t | \boldsymbol{\theta}_n)\right) \sum_{d=1}^{D} \boldsymbol{\theta}_{dn} \mathbf{C}_d \, h_\theta d\boldsymbol{\theta}_n = 0 \qquad (11)$$

$$\sum_{n=1}^{N} \int_{\boldsymbol{\theta}} \left(\mathbf{x}_n^t - E(\mathbf{z}^t | \boldsymbol{\theta}_n)\right) h_\theta d\boldsymbol{\theta}_n \mathbf{A} = 0 \qquad (12)$$

$$\sum_{n=1}^{N} \int_{\boldsymbol{\theta}} (\boldsymbol{\theta}_n - \boldsymbol{\gamma}\mathbf{W}_n)(\boldsymbol{\theta}_n - \boldsymbol{\gamma}\mathbf{W}_n)^t \, h_\theta d\boldsymbol{\theta}_n - N\boldsymbol{\Sigma} = 0 \qquad (13)$$

$$\boldsymbol{\gamma} = \left(\sum_{n=1}^{N} \int_{\boldsymbol{\theta}} \boldsymbol{\theta}_n h_\theta d\boldsymbol{\theta}_n \ \mathbf{W}_n^t \right) \left(\sum_{n=1}^{N} \mathbf{W}_n \mathbf{W}_n^t \right)^{-1} \quad (14)$$

Where

$$E\left(\mathbf{z}^t \mid \boldsymbol{\theta}_n\right) = \Psi \sum_{z\varepsilon\Omega} \mathbf{z} \exp\left\{ \mathbf{z}^T \left(\sum_{d=1}^{D} \boldsymbol{\theta}_{dn} \mathbf{C}_d \boldsymbol{\tau} + \mathbf{A}\boldsymbol{\varepsilon} \right) \right\} \quad (15)$$

These can solved iteratively using the approach of Bock and Aitkin (1981) and following the methods described in Adams and Wu (2007).

Fitting Standard Models

Control of the estimation of the scoring parameters is accomplished by use of the rasch, bock, and `scoresfree` options to the `model` command. For identification purposes (see below) the use of case constraints (`set constraint=cases)` is also required.

If the `rasch` option is used then a Rasch-type model is estimated and the scores are fixed at values defined through the `key` and `score` commands.

If the `scoresfree` option is used then a score is estimated for 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. This option provides a single score parameter for each item, so the model is equivalent to the generalised partial credit model (Muraki, 1992) applied to each generalised item.

If the `bock` option is used then a Bock nominal response model is estimated so 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.

In the case of dichotmous data the scoresfree and bock options are equivalent. For examples of their estimation (see Ockwell, 2015).

As with all ConQuest 4 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.

Fitting Models Specified Using Design Matrices

The design matrix **C** is used to specify the scoring parameter aspects of the model. The **C** matrix is generated automatically when the bock, and `scoresfree` options are added to the `model` command.

Greater flexibility can be obtained by importing a **C** matrix (see the `cmatrix` argument for the `import` command)

Identification Requirements

For Rasch models (models with fixed **C** 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 **C** 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.

References

Adams, R. J., & Macaskill, G., (2012). Score Estimation And Generalised Partial Credit Models. Retrieved from https://www.acer.edu.au/files/Conquest-Notes-6-ScoreEstimationAndGeneralisedPartialCreditModels.pdf.

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.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37, 29-51.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46, 443-459.

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

Ockwell, L. (2015). Modelling Polytomous Items with the Generalised Partial Credit and Bock Nominal Response Models. Retrieved from http://www.acer.edu.au/files/Tutorial-11-Generalised-Partial-Credit-and-Bock-Nominal-Response-Models.pdf.