Measurement and Research Department Reports 91-5

The Partial Credit Model With Non-Sequential Solution Strategies

N.D. Verhelst
H.H.F.M. Verstralen

Cito

3.4
91-5
95

The Partial Credit Model With Non-Sequential Solution Strategies

N.D. Verhelst
H.H.F.M. Verstalen

Cito Instituut voor Toetsontwikkeling
Bibliotheek

Cito
Arnhem, 1991

8501 016 1925

© Cito Arnhem
All rights reserved

Abstract

Although Masters (1982) used a step rationale to derive the partial credit model (PCM), it is argued in the present article that this interpretation is not implied by the model. It is shown that if, in the case of k Rasch-items, one observes only the number of items correct on K ($1<K<k$) subsets of items (which together partition the original k items), the PCM applies perfectly. More generally this implies that the content oriented interpretation of a formal model should be handled with care, and that the decisive criterion of acceptability of statistical models resides in statistical testing.

Key Words:Item Response Theory, Partial Credit Model.

Introduction

A model tailored to the analysis of item responses scored in ordered categories is Master's (1982) Partial Credit Model (PCM). To derive this model Masters uses the following item as an example:

$$\sqrt{\frac{0.75}{0.03} - 16} = ?$$

He assumes that the solution of this item has to proceed in three consecutive steps

1: .75/.03

2: 25 - 16

3: $\sqrt{9}$

His derivation implicitly assumes that a step cannot be successfully completed unless its predecessors have been. The partial score earned by a student for a particular item equals the number of steps he correctly completes. Together with the previous assumption a partial score j implies that exactly the first j steps are completed successfully.

This reasoning however, entails two problems. First the interpretation of the parameters. In the PCM there are three parameters associated with the example. Given the above reasoning it is a seductive conclusion that the parameter associated with a particular step does not change when it would be embedded in another item. Molenaar (1983) showed that this conclusion cannot be maintained, because the parameter value of a particular step depends on the parameter values for the other steps in the item. This interdependency implies that the parameter value of a step cannot be interpreted as a measure of its difficulty. Second, the more general interpretation of the PCM. It is certainly true that the PCM is appropriate for items that allow for a sequential solution as indicated by Masters. However, we would make a logical error if we concluded that item responses that fit the PCM, are necessarily solved in a sequential way and scored according to Master's rule. It cannot be known in advance that if, for instance, we would change the scoring rule of the item to a non-sequential type, the PCM does no longer fit. An obvious possibility is to credit every step yielding a formally correct solution, even if based on a faulty result from a previous step, with one point and to choose as the item score the ordinary sum score. With this rule a correct first and third step would add up to score 2, while Master's rule would only credit the first step with a score of 1. More in general, one could conjecture that the PCM is appropriate for all item scoring rules that result in ordered categories, without any assumption of a hierarchical solution strategy and its concomitant scoring rule. If Masters interpretation would be a necessary consequence of the validity of the PCM, one has to accept the fact that the model is appropriate only for a restricted type of items. The analysis of polytomous items which are clearly not necessarily sequential, would have to performed with other models (e.g., Bock & Aitkin, 1981). However, in most cases there is no way to decide whether an item is sequential or not. One could, of course, construct scoring rules that enforce a sequence interpretation. In the sequel we will show that this endeavour is in vain as the appropriateness of the PCM does not in any way necessitates a sequence interpretation. Section 2 contains a formal introduction to the partial credit model and section 3 treats a special case that clearly violates the assumption of a sequential solution, but, nevertheless, is in perfect agreement with the PCM. In section 4 the interpretation of the PCM parameters is discussed.

The Partial Credit Model

Let k be the number of items in a test and assume that an examinee may earn a score 0, 1, ..., m_i on item i, where the value m_i may be dependent on i. The item score (considered as a random variable) on item i is denoted by X_i. The PCM is defined by the following two axioms:

$$\mathrm{Prob}(X_i=j \mid \vartheta) \propto \exp(j\vartheta - \sum_{g=0}^{j} \beta^*_{ig}), \quad (j = 0, \ldots, m_i;\ i = 1, \ldots, k) \tag{1}$$

and

$$\mathrm{Prob}(X_i=j \text{ en } X_{i'}=j' \mid \vartheta) = \mathrm{Prob}(X_i=j \mid \vartheta)\,\mathrm{Prob}(X_{i'}=j'), \quad (i, i'=0, .., k;\ i \neq i') \tag{2}$$

Axiom (2) expresses local stochastic independence, a property of many IRT models. The parameter β^*_{ig} in (1) is associated with response category g. Note that (1) contains β^*_{i0} in the expression for every $j \in \{0, \ldots, m_i\}$. Because

$$\sum_{j=0}^{m_i} \text{Prob}(X_i=j \mid \vartheta) = 1,$$

(1) is equivalent with

$$\text{Prob}(X_i=j \mid \vartheta) = \frac{\exp[j\vartheta - \sum_{g=0}^{j} \beta^*_{ig}]}{\sum_{h=0}^{m_i} \exp[h\vartheta - \sum_{g=0}^{h} \beta^*_{ig}]}. \tag{3}$$

For a particular item i, the probabilities in (3) for all j ($j = 0, \ldots, m_i$) and all θ, determine the parameters β^*_{ig} ($g = 0, \ldots, m_i$) only up to an additive constant c_i. This can easily be shown by multiplying the numerator and denominator of the right part of (3) by $\exp(c_i)$. Defining

$$\beta_{ig} = \beta^*_{ig} - c_i ,$$

(3) can be rewritten as

$$\text{Prob}(X_i=j \mid \vartheta) = \frac{\exp[j\vartheta - \sum_{g=0}^{j} \beta_{ig}]}{\sum_{h=0}^{m_i} \exp[h\vartheta - \sum_{g=0}^{h} \beta_{ig}]}, \tag{4}$$

and it is clear that different values for c_i result in an identical set of probability functions for different sets of parameter values. A fruitful choice for c_i that simplifies the estimation calculations is:

$$c_i = \beta^*_{i0}, \quad (i = 1, \ldots, k),$$

from which it follows that

$$\beta_{i0} = 0, \quad (i = 1, \ldots, k),$$

and allows (4) to be rewritten as

$$\mathrm{Prob}(x_i=j\mid\vartheta) = \frac{\exp[j\vartheta - \sum_{g=1}^{j}\beta_{ig}]}{1+\sum_{h=1}^{m_i}\exp[h\vartheta - \sum_{g=1}^{h}\beta_{ig}]}. \tag{5}$$

Although the numerator in the right part of (5) suggests a sequence interpretation, this is definitely not imperative, as the following reparameterization (Glas, 1989, p. 14-15) shows.

Define

$$\eta_{ij} = \sum_{g=0}^{j}\beta_{ig}, \quad (j = 0, \ldots, m_i;\ i = 1, \ldots, k)$$

then (4) can be rewritten as

$$\mathrm{Prob}(X_i=j\mid\vartheta) = \frac{\exp[j\vartheta-\eta_{ij}]}{\sum_{h=0}^{m_i}\exp[h\vartheta-\eta_{ih}]}, \tag{6}$$

which is a slight generalization of a model proposed by Andersen (1977). Notice that, since the β's are unconstrained, so are the η's. Where the cumulative sum in the numerator of (4) could have suggested a sequence interpretation, this suggestion has disappeared in (6), where the parameters for the categories are arbitrary numbers, that do not impose a theoretically sequential solution strategy.

A Model for Testlets

The practice of testing often presents us with a set of items that share a common part, for example, a text to be read by the examinee. The text can cause its items to have higher intercorrelations than items associated with different texts. As a result simple IRT models such as the Rasch model must frequently be rejected as an adequate description of the item responses. The literature often refers to items with a common part with the term 'testlet'. Therefore, an appropriate analysis requires a model that adequately deals with the higher intercorrelations, or select a higher level model that directly addresses the testlet scores as the unit of analysis and not the item scores. As this does not determine yet the nature of these testlet scores, however, we can start with a particular definition of a testlet score and subsequently design a suitable model that adequately explains these scores. In this section we propose a modelled type of testlet analysis, that is, a test with k Rasch items partitioned into K (K>1) testlets. So, every testlet contains one or more items and the testlet score is the number of correctly solved items in the testlet. For every item we have

$$\mathrm{Prob}(X_i=1 \mid \vartheta) = \frac{\exp(\vartheta-\sigma_i)}{1+\exp(\vartheta-\sigma_i)}, \quad (i = 1, \ldots, k). \tag{7}$$

The partition in testlets can be denoted as follows. Let $I = \{1, 2, \ldots, k\}$, and consider K subsets of I, $A_\alpha \subset I$ ($\alpha = 1, \ldots, K$) with:

$$A_\alpha \cap A_\beta = \varnothing, \quad (\alpha \neq \beta),$$
$$\bigcup_\alpha A_\alpha = I,$$

and define

$$Y_\alpha = \sum_{i \in A_\alpha} X_i.$$

Suppose we want an analysis using only the testlet scores Y and not the item scores X. We derive the density function of Y using an example. Let $A_\alpha = \{1, 2, 3\}$, and define

$$\begin{aligned} \xi &= \exp(\vartheta), \\ \varepsilon_i &= \exp(-\sigma_i), \quad (i = 1, \ldots, k) \end{aligned} \tag{8}$$

From (7), and from the assumption of local stochastic independence it follows that

$$\mathrm{Prob}(Y_\alpha=2 \mid \xi) = \frac{\xi^2[\varepsilon_1\varepsilon_2+\varepsilon_1\varepsilon_3+\varepsilon_2\varepsilon_3]}{(1+\xi\varepsilon_1)(1+\xi\varepsilon_2)(1+\xi\varepsilon_3)}. \tag{9}$$

The sum between brackets in the numerator of (9) contains the products of exactly all possible pairs taken from three item parameters. Therefore, this sum is by definition equivalent to the elementary symmetric function (of the second order) of

$$\underline{\varepsilon}_\alpha = (\varepsilon_{i_1}, \ldots, \varepsilon_{i_p} \mid \{i_1, \ldots, i_p\} = A_\alpha).$$

Elementary symmetric functions of the order s will be denoted by $\gamma_s(\underline{\varepsilon}_\alpha)$. After expansion of the denominator of (9) and regrouping of terms, (9) can be rewritten as:

$$\mathrm{Prob}(Y_\alpha=2 \mid \xi) = \frac{\xi^2\gamma_2(\underline{\varepsilon}_\alpha)}{1+\xi\gamma_1(\underline{\varepsilon}_\alpha)+\xi^2\gamma_2(\underline{\varepsilon}_\alpha)+\xi^3\gamma_3(\underline{\varepsilon}_\alpha)}. \tag{10}$$

It is easily seen that this result can be readily generalized. Define $m_\alpha = |A_\alpha|$, then we have

$$\mathrm{Prob}(Y_\alpha=j \mid \xi) = \frac{\xi^j\gamma_j(\underline{\varepsilon}_\alpha)}{\sum_{h=0}^{m_\alpha}\xi^h\gamma_h(\underline{\varepsilon}_\alpha)}, \quad (j = 0, \ldots, m_\alpha). \tag{11}$$

Using (8) this can also be expressed by

$$\mathrm{Prob}(Y_\alpha=j \mid \vartheta) = \frac{\exp[j\vartheta+\ln\gamma_j(\underline{\varepsilon}_\alpha)]}{\sum_{h=0}^{m_\alpha}\exp[h\vartheta+\ln\gamma_h(\underline{\varepsilon}_\alpha)]}, \quad (j = 0, \ldots, m_\alpha). \tag{12}$$

Now define

$$\eta_{\alpha j} = -\ln\gamma_j(\underline{\varepsilon}_\alpha), \quad (j = 0, \ldots, m_\alpha), \tag{13}$$

then the formal equivalence of (6) and (9) is evident. Thus, by considering 'Rasch' testlets as multi-categorical items, these testlets are perfectly modeled by the PCM. Moreover, the item scores can definitely not be considered as resulting from a sequential solution process. The items in a testlet are not subject to any hierarchy as is implied by the Rasch model, and the assumption therefore that step j+1 can only be successfully completed conditional on the success of step j does not hold.

The Interpretation of the Parameters.

When we try to offer an interpretation for the parameters in the special case of the 'Rasch' testlets, we should be concerned not to make the same mistake for which we try to warn here. The next statement is surely not true: 'If testlet scores, defined as the number correct score of the items in the testlet, can be adequately described by the PCM, then for every testlet its set of items can be considered a set of Rasch items. The parameters associated with these items can be found by the inverse transformation of (13), that is, with given η's to solve (13) for the ϵ's.' The nonvalidity of this statement easily follows from the fact that the η's are not constrained, whereas it is certainly not true that for m_α arbitrary numbers (13) can be solved. However, even if (13) has a solution, then it is still not inevitable to decide for a hierarchical interpretation; at most this may be called clever, but it is not connected with the PCM as such. Moreover, note that if (13) has a solution for the true η's, this does not necessarily apply to the estimated η's. One can estimate the η's under the restriction (13), but this means that one does not assume the general PCM to hold, because the full parameter space is not used.

More in general it can be stated that no interpretation that hints at behavioral or cognitive strategies is indicated by the model parameters or can be compellingly associated with them. The direction of interpretation towards the model can be useful and clarifying; the other way around (from model to interpretation) is never compelling and in general unnecessary, and perhaps, harmfully constraining.

A totally different type of interpretation can be based on the model itself. However, this type of interpretation can barely be expected to transcend a paraphrasing of the model in a truly unveiling way. For example, with the PCM the following 'interpretation' is always true: the parameter β_{ij} (under the constraint (7)) is the unique value of θ with the same expected probability for score j on item i as score j-1 (j>0) (Glas, 1989). However, this interpretation is simply implied by the unique solution for θ of the equation

$$\mathrm{Prob}(Y_i=j \mid \vartheta) = \mathrm{Prob}(Y_i=j-1 \mid \vartheta),$$

which yields $\theta = \beta_{ij}$.

And, of course, more interpretations can be found in a similar way that may be useful to comprehend the PCM. For example, it can be shown that, if $\beta_{ij} > \beta_{i,j-1}$ ($j>1$), for all θ in the open interval (β_{ij}, $\beta_{i,j-1}$), score j-1 is the modal score (that is, it is the score with the highest probability). If the analysis shows that this monotonicity is clearly violated, then one or more scores will never be modal. In that case one could consider another scoring rule.

Conclusion

The above discussion is, of course, not meant to suggest that behavioral interpretations are to be avoided or cannot be fruitful. Nor was it written to advocate the PCM as <u>the</u> formal model of choice for all data that allow an interpretation as a set of ordered categories. The model can better be looked at as a formal restriction on the response probabilities, with the status of a statistical hypothesis. The 'analysis according to the model' then naturally divides in an estimation part and a testing part. The estimation part can be ignored as a technical matter that can be dealt with in an algorithmic fashion. The testing part, however, offers the opportunity to test the acceptability of behavioral or cognitive interpretations. Of course this does not mean that this testing is easy. If, for example, one estimates the PCM-parameters under the restriction (13), the resulting parameter space is not a proper subspace of the unrestricted one - the restrictions can be written as a set of inequalities - and in both cases the parameter space has the same dimensionality. So the common likelihood ratio test does not apply, and it is not quite clear how the restricted model might be tested against the unrestricted one.

As a general conclusion, there is no evidence that the PCM should be restricted to items which allow for a step interpretation. The special case elaborated in the previous paragraphs indicates that the model offers a perfect description in a case where this interpretation is explicitly excluded. So the PCM may be considered as a candidate model for data where the scores can be interpreted as ordered categories and its adequacy should not primarily be judged upon by interpretation but by statistical testing.

References

Andersen, E.B. (1977). Sufficient statistics and latent trait models. *Psychometrika, 42*, 69-81.

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

Glas, C.A.W. (1989). *Contributions to estimating and testing Rasch Models*. Arnhem, Cito.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Molenaar, I.W. (1983). Some improved diagnostics for the Rasch model. *Psychometrika, 48*, 49-72.

Recent Measurement and Research Department Reports:

91-1 N.D. Verhelst & N.H. Veldhuijzen. A New Algorithm For Computing Elementary Symmetric Functions And Their First And Second Derivatives.

91-2 C.A.W. Glas. Testing Rasch Models For Polytomous Items: With An Example Concerning Detection Of Item Bias.

91-3 C.A.W. Glas & N.D. Verhelst. Using The Rasch Model For Dichotomous Data For Analyzing Polytomous Responses.

91-4 N.D. Verhelst & C.A.W. Glas. A Dynamic Generalization Of The Rasch Model.