Measurement and Research Department Reports 91-4

A Dynamic Generalization Of The Rasch Model

N.D. Verhelst
C.A.W. Glas

Cito

A Dynamic Generalization Of The Rasch Model

N.D. Verhelst
C.A.W. Glas

Cito
Arnhem, 1991

Abstract

By combining the common Rasch model with the concept of incomplete designs, the former can be made suited as a dynamic model: items in a test are considered to consists of a collection of conceptual items, one of which is presented to the testee dependent on his preceding responses and/or the feedback (s)he got. It is shown that marginal maximum likelihood (MML) is suited as an estimation procedure, and the conditions are specified where conditional maximum likelihood estimates (CML) may be obtained. A hierarchic family of dynamic models is presented, and it is shown how to test special cases against more general ones. Furthermore, it is shown that the model presented here is a generalization of a class of mathematical learning models, known as Luce's beta-model.

Key words: Rasch model, missing data, incomplete designs, dynamic models, mathematical learning theory.

Introduction

Perhaps the most outstanding feature of IRT models is the fact that interindividual differences in behaviour are explained at the model level, implying that subjects need not be considered as statistical replicates of each other nor that it is necessary to sample multiple observations from the same subject under assumedly constant conditions. The way this is accomplished is by introducing so-called incidental parameters, one parameter associated with each individual. Although these parameters can be treated formally as any other parameter in the model, a major problem is associated with their presence in parameter estimation. Maximizing the likelihood for example will in general not yield consistent estimates of the incidental parameters and of the other (sometimes called structural) parameters of the model (Neyman & Scott, 1948). There are several ways to cope with this problem. Two estimation methods are rather popular in IRT. In the first method, the sufficient statistics for the incidental parameters (if they exist) are considered as constant, and the conditional likelihood, given these statistics, is maximized. This conditional likelihood is by definition independent of the incidental parameters and Andersen (1973) proved that this method yields consistent estimates of the structural parameters. This method is commonly labelled as Conditional Maximum Likelihood (CML). In the other approach the incidental parameters are no longer considered as fixed constants, but are treated as realizations of a (non-observed or latent) random variable, whose distribution is assumed to belong to a certain family of distributions. In many applications this family is parametrized by a finite number of parameters (for example the family of normal distributions). Since the latent variable is not observed, it is integrated out from the likelihood function, yielding the so called Marginal Likelihood function. This function is maximized with respect to the structural model parameters and the parameters of the distribution jointly, yielding the Marginal Maximum Likelihood (MML) estimates, which by a seminal paper of Kiefer & Wolfowitz (1956) are proved to be consistent under very mild regularity conditions. The use of CML is restricted to models which have minimal sufficient statistics for the incidental parameters, and among the IRT models commonly applied very few are in this class. (See Verhelst & Eggen (1989) for a general characterization.) The most famous example is the Rasch model, and both estimation methods have been successfully applied to this model. (For CML, see Rasch (1960) and Fischer (1974), for MML, Thissen (1982) and Glas (1989).)

Several authors (Fischer, 1981, Mislevy, 1984) have pointed at the fact that it is possible to obtain consistent estimates of the structural parameters in the Rasch model, from partially incomplete data, or data collected in an incomplete design. It is the purpose of the present paper to investigate the applicability of the Rasch model in a dynamic context by manipulating the missing data concept on a set of seemingly complete data. This manipulation is convenient to get around the restrictions implied by the central axiom of local stochastic independence common to most latent trait models. For one thing, this axiom implies that a correct response by a given subject is independent of his responses on previous items or trials, thus seemingly excluding the so-called subject controlled learning models where the probability of a correct answer depends explicitly on the response pattern given thus far (Sternberg, 1963). Although there exist generalizations of the Rasch model, where this dependence can be explicitly modeled (Kelderman, 1984; Jannarone, 1986), parameter estimation is difficult in these models, and is restricted to cases with a rather restricted number of parameters. The approach used in the present paper combines the Rasch model with the missing data concept and with linear restrictions on the parameter space, yielding a wide range of dynamic models which coincide with the Rasch model as originally defined. The basic approach consists in conceiving an item (or trial in a learning experiment) as a collection of 'conceptual' or 'technical' items , one of which is supposed to be administered to each subject, depending, for example, on the pattern of previous responses. This is possible because of the basic symmetry in the Rasch model, where a change in the latent ability can be equally well be conceived as a complementary change in the item difficulty. The details of this approach are the subject matter of the next section.

The idea presented here is not new ; in fact it has been considered by Fischer (1972, 1983). Working in the context of conditional maximum likelihood estimation, Fischer rejected this approach because the model thus constructed was not estimable. In the section 'estimation', it will be shown that the models constructed are in general estimable under MML, and that for some special subclass CML applies. A separate short section is devoted to statistical tests of special cases of the model against more general ones. Finally it will be shown that a class of mathematical learning models, known as Luce's beta model and generalizations thereof, are a special case of our model.

The Dynamic Rasch Model

In mathematical learning theory (see Sternberg (1963) for a general introduction) the control of change in the behaviour is attributed generally to two classes of events: one is the behaviour of the responding subject itself, the other comprises all events which occur independently of the subject's behaviour, but which are assumed to change that behaviour. Models which only allow for the former class are called 'subject controlled', if only external control is allowed, the model is 'experimenter controlled', and models were both kinds of control are allowed are labelled 'mixed models'. As an example of 'experimenter control' assume that during test taking, the correct answer is given to the respondent after each item response. If it is assumed that learning takes place under the influence of the feedback (generally referred to as reinforcement) independent of the correctness of the given response, the model is experimenter controlled; if it is assumed, however, that learning takes place only if the subject gave a wrong answer, the model is mixed. In the sequel it will be assumed that all controlling events can be binary coded, that the subject control can be modeled through the correctness of his responses on past items, and that experimenter control expresses itself at the level of the item. Let $X = (X_1,..,X_i,..,X_k)$ be the vector of response variables, taking values 1 for a correct answer and zero otherwise, and let $Z = (Z_1,..,Z_i,..,Z_k)$ be the binary vector representing the reinforcement event occurring after the response has been given. The value Z_i takes is assumed to be independent of X. The prototype of a situation where this independence is realized is the so called prediction experiment where the subject has to predict a sequence of independent Bernouilli events, indexed by a constant parameter π. (for example the prediction of the outcome of coin tosses with a biased coin). The outcome itself is independent of the prediction and it is observed frequently that in the long run the relative frequency of predicting each outcome matches approximately the objective probability of the outcomes (Estes, 1972). It can be assumed therefore that the outcome itself acts as a reinforcer and is the main determinant of the change in the behaviour. Notice, that in the example given, Z_i is a random variable which is independent of X. Therefore any model which assumes control over the responses only through the outcomes is experimenter controlled.

Define the partial response vector X^i $(i>1)$ as

$$X^i = (X_1,..,X_{i-1}) \tag{1a}$$

and the partial reinforcement vector Z^i $(i>1)$ as

$$z^i = (z_1,..,z_{i-1}). \tag{1b}$$

The model that will be considered is in its most general form given by

$$\text{Prob}(X_i=1|\vartheta,\delta_i,x^i,z^i) = \frac{\exp[\vartheta+\delta_i+f_i(x^i)+g_i(z^i)]}{1+\exp[\vartheta+\delta_i+f_i(x^i)+g_i(z^i)]}, \tag{2}$$

where x^i and z^i represent realizations of X^i and Z^i respectively, and $f_i(.)$ and $g_i(.)$ are arbitrary real-valued functions. Since the domain of these functions is discrete and finite, the possible function values can be conceived of as parameters. It is clear that the general model is not identified, because the number of parameters outgrows by far the number of possible response patterns. So some suitable restrictions will have to be imposed.

A general restriction, which has frequently been applied in mathematical learning theory is to require that the functions f_i and/or g_i are symmetric in their arguments, yielding models with commutative operators. Since the arguments are binary vectors, symmetry implies that the domain of the function can be restricted to the sum of the elements in the vectors X^i and Z^i respectively. Defining the variables R_i and S_i as

$$R_i = \begin{cases} \sum_{j=1}^{i-1} X_j, & (i>1), \\ 0, & (i=1), \end{cases} \tag{3a}$$

and

$$S_i = \begin{cases} \sum_{j=1}^{i-1} Z_j, & (i>1), \\ 0, & (i=1), \end{cases} \tag{3b}$$

with realizations r_i and s_i respectively, and assuming symmetry of the functions g_i and f_i, (2) reduces to

$$\text{Prob}(X_i=1|\vartheta,r_i,s_i) = \frac{\exp[\vartheta+\delta_i+\beta_i(r_i)+\gamma_i(s_i)]}{1+\exp[\vartheta+\delta_i+\beta_i(r_i)+\gamma_i(s_i)]}, \tag{4}$$

where $\beta_i(0)$ and $\gamma_i(0)$ are defined to be zero for all i. If all β and γ are equal zero, no transfer takes place, and (4) reduces to the common Rasch model; if all β's are zero, and at least one of the γ's is not, the resulting model is experimenter controlled, if all γ's are zero and at least one of the β's is not, the model is subject controlled, and in the other cases a mixed model results. Notice that (4) implies that no forgetting occurs: the influence of an equal number of correct responses and/or an equal number of positive reinforcements has the same influence on the behaviour, immaterial of their temporal distance to the actual response. This somewhat unrealistic assumption is the price to be paid for the elegant mathematical implications of commutative operators. In the sequel however, a model will be discussed where this symmetry is at least partially abandoned.

In order to see how (1) fits in the ordinary Rasch model using the concept of incomplete data, only the subject controlled subcase of (4) will be considered in detail, i.e. it will be assumed that all γ's are identical zero. Notice that it is assumed that the transfer that takes place does not depend on the initial ability ϑ of the respondent, and that any change in the ability (the increment $\beta_i(r_i)$) can be translated in a change in the difficulty of the items. So the difficulty of an item is conceived of as composed of an 'intrinsic' parameter δ_i and a dynamic component $\beta_i(r_i)$. The latter is not constant, but depends on the rank number of the item in the sequence (hence the subscripted r), as well as on the specific capability of the item to induce learning effects (hence the subscripted β). These two effects can in principle be disentangled by experimental design, but in order not to overcomplicate the model, it will be assumed that the order of presentation is the same for all subjects.

Let a 'real' item i be associated with a collection of 'conceptual' items, denoted by the ordered pair (i,j), j=0,..,i-1. The conceptual item is presented to and responded by all subjects who gave exactly j correct responses to the i-1 preceding 'real' items. Associated with response pattern X is a design vector D(X), with elements $D(X)_{ij}$, (i=1,..,k; j=0,..i-1) defined by

$$D(X)_{ij} = \begin{cases} 1 \text{ if } R_i=j \\ 0 \text{ otherwise.} \end{cases} \tag{5}$$

Response pattern X is transformed into a response pattern Y(X) with elements $Y(X)_{ij}$ (i=1,..,k;j=0,..,i-1) using

$$Y(X)_{ij} = \begin{cases} 1 \text{ if } D(X)_{ij}=1 \text{ and } X_i=1, \\ 0 \text{ if } D(X)_{ij}=1 \text{ and } X_i=0, \\ c \text{ if } D(X)_{ij}=0, \end{cases} \tag{6}$$

where c is an arbitrary constant. Table 1 displays the set of possible response patterns X and the associated vectors Y(X) for k=3. (The constant c is replaced by a star.)

TABLE 1

The transformation from real to conceptual items for k=3

real items			conceptual items						sum score
1	2	3	(1,0)	(2,0)	(2,1)	(3,0)	(3,1)	(3,2)	
1	1	1	1	*	1	*	*	1	3
1	1	0	1	*	1	*	*	0	2
1	0	1	1	*	0	*	1	*	2
1	0	0	1	*	0	*	0	*	1
0	1	1	0	1	*	*	1	*	2
0	1	0	0	1	*	*	0	*	1
0	0	1	0	0	*	1	*	*	1
0	0	0	0	0	*	0	*	*	0

The probability of an observed response pattern y(x) jointly with the observed design vector d(x) can now be given by

$$\text{Prob}(y(x), d(x) \mid \vartheta, \xi) = \frac{\exp[\sum_{i,j} y(x)_{ij} d(x)_{ij} (\vartheta + \xi_{ij})]}{\prod_{i,j} [1 + \exp(\vartheta + \xi_{ij}]^{d(x)_{ij}}}, \qquad (7)$$

with ξ a k(k+1)/2 dimensional vector with elements ξ_{ij} (i=1,..,k;j=0,..,i-1) where $\xi_{ij} = \delta_i + \beta_i(j)$. It is easily verified that this is nothing but a simple reparametrization of the original model. Equation (7) is a simple generalization of the likelihood function for the Rasch model to incomplete designs. Of course, a similar derivation may be made for experimenter controlled models, and a straightforward generalization yields a similar result for mixed models. Although the likelihood function for the experimenter controlled case is formally identical to (7), with the design vector being a function of Z instead of X, the estimation problems are quite different, as will be discussed in the next section.

Estimation

Glas (1988) has investigated the estimation problems in the Rasch model for so called multi-stage testing procedures, i.e., designs where the sequence of tests administered is controlled by the sequence of test scores of the testee. On the level of conceptual items, the design used to analyze data under a subject controlled model in the present paper can be viewed as a limiting case of a multi-stage testing design, where all tests consists of one item only, and the next test to be administered depends on the sum score R_i obtained in the preceding tests. The main result of Glas is the conclusion that in the case of a multi-stage design, CML estimation is not applicable, while MML in general yields consistent estimates.

With respect to CML estimation, the argument can be summarized as follows. The test administration design and the sum score are a sufficient statistic for ϑ. CML estimation in incomplete designs amounts to maximize the conditional likelihood function, where the condition is the sum score and the design jointly. (See Fischer, 1981 for details). For the design presented in Table 1, it is easily verified that, given the design (i.e., the location of the stars) and the sum score, the response vector X is completely determined. This means that the restricted sample space, as a result of the condition, is a singleton, and its likelihood is trivially identical 1, so that it cannot be used to estimate the structural parameters. The situation is totally different for experimenter controlled models, where the control over the design vector is completely independent of X, so that for every observed X, k! different transformations Y(X) are possible. So, conditioning on the design implies no restriction whatsoever on the sample space of X and CML can be applied straight on.

In order to be able to estimate the parameters in subject controlled or mixed models, one has to take recourse to the MML procedure. In contrast to complete designs however, one has to use (7) as the likelihood function, implying that not only the distribution of Y(X) has to be taken into account but also the distribution of D(X), which is also dependent on the model parameters. However, by a theorem of Rubin (1976), it can be shown that in the present case the design is ignorable, meaning that maximum likelihood estimates yield the same values if the design variables are treated as fixed constants. The core of the argument amounts to the statement that the subject control over the design is completely reflected in the observed responses, or conversely, the distribution of D(X) is independent of the value of the non-observed responses (i.e., what is hidden behind the stars in Table 1).
Now let n_x be the number of observations having X=x, then the log-likelihood function to be maximized is given by

$$\ln L(\xi,\varphi) = \sum_x n_x \ln \int \text{Prob}(y(x)\,|d(x),\vartheta;\xi)\, g(\vartheta;\varphi)\,d\vartheta, \qquad (8)$$

where $g(\vartheta;\varphi)$ is the probability density function of ϑ, indexed by the parameter vector φ. Notice that in (8), the design vector appears as a condition, as opposed to (7) where it appears jointly with the response vector. This change is justified by Rubin's argument on the ignorability of the design. If it is assumed that ϑ is normally distributed, (8) has the same form as a usual item response log-likelihood function with incomplete data (see e.g. Mislevy, 1984), and as a result, the computational procedures for estimating the model, such as the EM-algorithm (Bock & Aitkin, 1981) can be directly applied. With respect to identification of the model, some assumption has to be made with respect to the distribution: if it is assumed that ϑ is normally distributed with mean zero and variance σ^2, the parameters ξ need no further restrictions. Because of the equivalence of the dynamic model with the Rasch model, as applied in incomplete designs, the expressions for the likelihood equations and the asymptotic confidence intervals are equally valid in both cases and can for instance be found in Glas & Verhelst (1989) or Glas (1989). The conditions for the existence of a solution to the likelihood equations with MML are not yet completely clarified, although experience with the model shows that few restrictions seem to exist. At any rate it is necessary that at least one positive and one negative response is given to each conceptual item. Although this restriction is not very severe when working in complete designs, it should be kept in mind that the number of conceptual items grows quadratically with the number of real items: for k real items there are (in the subject controlled version of (4)) $k(k+1)/2$ conceptual items. This implies that for the parameters to be estimable, substantial variability in response patterns will be required. Of course it is always possible to remove conceptual items ad hoc from the analysis, but a more elegant procedure consists in imposing restrictions on the parameter space of the general model and thus generating special cases. This is the topic of the next section.

Linear Restrictions on the Parameter Space

Starting with the model defined by (7) and (8), it is possible to derive several interesting special cases by imposing linear restrictions on the parameters ξ. So let η be an m-dimensional vector, $m < k(k+1)/2$, such that $\eta = B\xi$, with B a (constant) matrix of rank m. As above it is assumed that ϑ is normally distributed with mean zero and variance σ^2. The following models are identified.

(i) Amount of learning is independent of the preceding items, yielding the restrictions

$$\xi_{ij} = \begin{cases} \delta_i & \text{if } j=0. \\ \delta_i+\beta_j & \text{if } j>0. \end{cases} \tag{9}$$

(ii) As a further restriction of the previous model, one may suppose that the amount of learning after each success is constant, yielding

$$\xi_{ij} = \begin{cases} \delta_i & \text{if } j=0. \\ \delta_i+j\beta & \text{if } j>0. \end{cases} \tag{10}$$

(iii) A two operator model can be constructed by assuming that the amount of change in latent ability is not only a function of the number of previous successes, but also of previous failures. The most general version of a two operator model with commutative operators is given by

$$\xi_{ij} = \begin{cases} \delta_i & \text{if } i=0. \\ \delta_i+\beta_j & \text{if } i>1 \text{ and } j=i-1, \\ \delta_i+\varepsilon_{i-j-1} & \text{if } i>1 \text{ and } j=0, \\ \delta_i+\beta_j+\varepsilon_{i-j-1} & \text{otherwise.} \end{cases} \tag{11}$$

It can easily be checked that (11) is a full rank reparametrization.

(iv) Analogously to (i), one can assume that the amount of transfer is independent of the item, specialising case (iii) further by imposing $\beta_j = j\beta$ and $\varepsilon_j = j\varepsilon$.

(v) It can be assumed that the effect of an incorrect response is just the opposite of the effect of a correct response, by imposing the extra restriction $\delta = -\varepsilon$ on the model defined in (iv).

(vi) Finally, one could assume as a kind of limiting case that the amount of learning is the same irrespective of the correctness of the preceding responses. Formally this can be modeled as a further restriction on case (iv) by putting $\beta = \varepsilon$. However in this case the model is no longer identified, if the rank order of presentation is the same for each respondent, because the parameter of each conceptual item (i,j) is given by $\delta_i + (i-1)\beta$, and the value of β can be freely chosen, because adding an arbitrary constant c to β can be compensated for by subtracting c/(1-i) (i>1) from δ_i. Besides, in this case, there is no more subject control. If for example the learning effect is caused by giving feedback after each response, or by the item text itself, or in generic terms by some reinforcer not under control of the testee, the above model is also a limiting case of experimenter control, limiting in the sense that there is no variability in the reinforcement schedule. So the solution to the identification problem is simple: introducing variability in the reinforcement schedule will solve the problem. For this restricted model, where the amount of learning increases linearly with the rank number of the item in the sequence, it suffices to administer the test in two different orders of presentation to two equivalent samples of subjects. Let in the ordered pair (i,j) (the conceptual item) i represent the identification label of the item and j the rank number in the presentation, then $\xi_{ij} = \delta_i + (j-1)\beta$, and the model is identified if there is at least one i such that (i,j) and (i,j'), $j \neq j'$ are conceptual items.

Testing the Model

In many instances, one may be interested in testing the validity of a more restricted model against a more general case. Let $\eta_2 = B_2\xi$ represent the general case and let $\eta_1 = B\eta_2 = BB_2\xi$ with $m_1 = \text{rank}(B) < \text{rank}(B_2) = m_2$ represent the restricted model. Furthermore let $L_1(\hat{\eta}_1, \hat{\sigma}^2{}_1)$ and $L_2(\hat{\eta}_2, \hat{\sigma}^2{}_2)$ stand for the maximum of the likelihood function in the restricted and the general model respectively. Since both models are parametrized multinomial models, standard asymptotic theory applies and

$$\lambda = -2\ln\frac{L_1(\hat{\eta}_1, \hat{\sigma}_1)}{L_2(\hat{\eta}_2, \hat{\sigma}_2)} \qquad (12)$$

has an asymptotic chi square distribution with $m_2 - m_1$ degrees of freedom.

The Relationship With Mathematical Learning Theory

Mathematical learning theory is an area that has kept much attention among mathematical psychologists in the early sixties. The chapters 9, 10 and 17 of the Handbook of Mathematical Psychology (Luce, Bush & Galanter, 1963) were entirely devoted to formal learning models and contain many interesting results. To get a good impression of the scope of these models, and of the problems that were recognized to be difficult, an example will be given. In the avoidance training, an animal is placed in a box, where it can avoid an electric shock by jumping over a barrier within 10 seconds after the occurrence of a conditioned stimulus (a buzzer sound). In a simple learning model it is assumed that (a) learning, i.e., a change in the tendency to avoid the threatening situation before being shocked, occurs only on escape trials, i.e. when the animal escapes after being shocked; (b) the 'inherent' difficulty of the situation is constant and (c) there are no initial differences between the animals in the initial tendency to avoid chocks. Of course this theory implies subject control. If the trials are identified with 'real items', then (b) and (c) imply that δ_i is constant, say $\delta_i = \delta$, and that the initial ability ϑ is constant. Letting an escape be a success, the probability of a success on trial i, given r_i successes on previous trials is given by

$$\mathrm{Prob}(X_i=1|\nu,R(X)=j) = \frac{\nu\alpha^j}{1+\nu\alpha^j}, \tag{13}$$

where $\nu = \exp(\vartheta+\delta)$ and $\alpha = \exp(\beta)$, which is known as Luce's (1959) one-operator beta model. If it is assumed that there is some learning following a failure (an avoidance), then

$$\mathrm{Prob}(X_i=1|\nu,R(X)=j) = \frac{\nu\,\alpha_1^j\,\alpha_2^{i-j-1}}{1+\nu\,\alpha_1^j\,\alpha_2^{i-j-1}}, \tag{14}$$

with $\alpha_1 = \exp(\beta)$ and $\alpha_2 = \exp(\varepsilon)$, which is equivalent to Luce's two-operator beta-model. So this model is just a special case of case (iv) discussed above.

The assumptions of no variability in the difficulty parameters or in the initial ϑ are characteristic for the many learning models developed roughly between 1955 and 1970. The lack of variability in the difficulty parameters may be attributed mainly to the fact that most applications concerned experiments with constant conditions over trials, while the assumed constancy in initial ability was recognized as a problem: '(..)in most applications of learning models, it is assumed that the same values of the initial probability (...) characterize all the subjects in an experimental group. (...) It is convenience, not theory, that leads to the homogeneity assumption' (Sternberg, 1963, p. 99). The convenience has to be understood as the lack of tools at that time to incorporate individual differences as a genuine model component, and the rather crude estimation procedures which were used. Maximum likelihood methods were used, although rarely, only in conjunction with Luce's beta-model; but this model was by far less popular than the family of linear models introduced by Bush & Mosteller (1951), where the probability of a success can be expressed as a linear difference equation, while Luce's model can be written as a linear difference equation in the logit of success probability. Most estimation methods in the linear model were modified moment methods, frequently yielding problems, because it was clearly acknowledged that suppression of interindividual variability would underestimate the variance of almost every statistic: '(...) unless we are interested specifically in testing the homogeneity assumption, it is probably unwise to use an observed variance as a statistic for estimation, and this is seldom done' (Sternberg, ibid.).

The final example that will be given serves a triple purpose: (i) it is an example of a mixed model with non commuting operators, i.e., (1) applies but (4) does not; (ii) it illustrates the rich imagination of the older learning theorists and at the same time their struggle to handle rather complex mathematical equations and (iii) it yields a nice suggestion to construct a statistical test for the axiom of local stochastic independence in the Rasch model. The model is the 'logistic' variant of the one-trial perseveration model of Sternberg (1959). The model was developed because the autocorrelation of lag 1 in the response pattern X was larger than predicted by the theory of the one operator linear model, suggesting that apparently there was a tendency to repeat the preceding response. (the experiment was a binary choice experiment, where one choice was systematically rewarded by a food pellet; the subjects were rats). Defining the choice of the non-rewarded response as a success, the model Sternberg proposed is given by

$$p_i = (1-b)a^{i-1}p_1 + bX_{n-1}, \quad (i \geq 2,\ 0<a,b<1), \tag{15}$$

where p_i = Prob(X_i=1), a is the parameter expressing the learning rate, and b is the perseveration parameter, expressing the tendency to repeat the previous response. The logistic analogue, mentioned in Sternberg (1963, p.36), but not analyzed is given by

$$\text{logit}\, p_i = \vartheta + (i-1)\beta + \gamma X_{i-1}, \quad (i>1), \tag{16}$$

where ϑ = logit p_1 is treated as a constant. Equation (16) is readily recognized as a special case of (2), where δ_i = 0, $g_i(Z_i)$ = $(i-1)\beta$ and $f_i(X^i)$ = γX_{i-1}. Notice that $f_i(.)$ is not symmetrical, so (16) is not a special case of (4). Notice further that (16) is more flexible than (15): by the restrictions put on the perseveration parameter b, tendencies to alternate the response require another model, while in (16) a positive γ expresses a perseveration tendency, while a negative γ expresses a tendency to alternate.

It is immediately clear that (16) violates the assumption of local stochastic independence. Now suppose data collected with an attitude questionnaire are to be analyzed with the common Rasch model, but there is some suspicion of response tendencies, in the sense of for example a tendency to alternate responses. Model (16) is readily adapted to this situation: set $\beta = 0$, allow variation in the 'difficulty' parameters δ_i and in the latent variable ϑ. There are 2k-1 conceptual items: (i,0), (i,1) for i>1, and (1,1) ≡ (1,0), the second member of the ordered pairs being equal to the previous response. The assumption of local stochastic independence can be tested with a likelihood ratio test as explained in the previous section using a restricted model where $\gamma = 0$, but this is nothing else than the common Rasch model.

Conclusion

The model proposed in this paper (equation (2)) is in fact nothing else than the common Rasch model. Its power to handle dynamic data stems from the manipulation with incomplete designs by the introduction of 'conceptual items'. The general model, defined by (2), however, is much too general to be identified, and therefore a large subclass of models is defined, where the influence of past events, whether subject controlled or experimenter controlled, does not extinguish. Although this family is rich, it may be rejected as a whole because of this insensitivity, in spite of the attractive features of commutative operators. However, specializations of (2) to non commuting subfamilies, such as the one trial perseveration model are feasible, and open up possibilities to test the basic axiom of local stochastic independence. Model tests are straightforward in a class of nested submodels: simple likelihood ratio tests are applicable. As to the parameter estimation, two results are crucial in the case of subject controlled or mixed models. The first is the clear exposition of the reasons why CML is not applicable (Glas, 1988), so that recourse had to be taken to MML. The second is the important result by Rubin (1976), which allows to ignore the stochastic character of the design variable. Although no Bayesian approach was used for the estimation, it may be reconforting to learn that the design is also ignorable in this context, as was shown by Rubin in the same article. CML is only applicable in experimenter controlled models. As a possible application, it was shown how to develop a simple data collection design to test for order effects in test taking.

The comparison with a large class of mathematical learning theories, viz. Luce's beta-model, and all 'logistic' variants, such as (16), which does not follow from Luce's choice axiom, revealed that these learning models are all special cases of (2) in two respects: they do not allow variability in the 'inherent' difficulty of the item or trial and most important, they are not able to explain individual differences. At the time mathematical learning theories where flourishing, important papers, nowadays acknowledged as the founding papers of IRT (Ferguson, 1942; Lawley, 1943; Lord, 1952 and Rasch, 1960), were available. The contact seemingly never took place. The model presented here does not have the two aforementioned restrictions: the variability of the latent variable is almost the defining characteristic of IRT models and by allowing items or trials to have their own specific parameters, learning theories can find their applications outside the psychological laboratories, for example, in educational testing.

Not all problems are solved however. A very hard one was hidden in an ellipsis in the first quotation of Sternberg. The full text reads: '... in most applications of learning models, it is assumed that the same values of the initial probability and other parameters characterize all the subjects in an experimental group...'. The underscored text refers to the possibility that learning effects also may show individual differences. These, however, are not easily built into the general model, this means it is easy to attach a subscript to some parameter symbols, but it is not easy to estimate individual learning parameters.

References

Andersen, E.B. (1973). *Conditional Inference and Models for Measuring.* (Unpublished dissertation). Copenhagen, Mentalhygienisk Foskningsinstitut.

Bock, R.D. & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: an Application of an EM-algorithm. *Psychometrika*, 46, 443-459.

Bush, R.R. & Mosteller, F. (1951). A Mathematical Model for Simple Learning. *Psychological Review*, 58, 313-323.

Estes, W.K. (1972). Research and Theory on the Learning of Probabilities. *Journal of the American Statistical Association*, 67, 81-102.

Ferguson, G.A. (1942). Item Selection by the Constant Process. *Psychometrika*, 7, 19-29.

Fischer, G.H. (1972). A Step towards Dynamic Test Theory. Research Bulletin nr. 24. Institute of Psychology, University of Vienna.

Fischer, G.H. (1974). *Einführung in die Theorie Psychologischer Tests*. Bern, Huber.

Fischer, G.H. (1981). On the Existence and Uniqueness of Maximum Likelihood Estimates in the Rasch Model. *Psychometrika*, 46, 59-77.

Fischer, G.H. (1983). Logistic Latent Trait Models with Linear Constraints. *Psychometrika*, 48, 3-26.

Glas, C.A.W. (1988). The Rasch Model and Multi-stage Testing. *Journal of Educational Statistics*, 13, 45-52.

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

Glas, C.A.W. & Verhelst, N.D. (1989). Extensions of the Partial Credit Model. *Psychometrika*, 54, 635-659.

Jannarone, R.J. (1986). Conjunctive Item Response Theory Kernels. *Psychometrika*, 51, 357-373.

Kelderman, H. (1984). Loglinear Rasch Model Tests. *Psychometrika*, 49, 223-245.

Kiefer, J. & Wolfowitz, J. (1956). Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. *Annals of Mathematical Statistics*, 27, 887-903.

Lawley, D.N. (1943). On Problems Connected with Item Selection and Test Construction. *Recordings of the Royal Society of Edinburgh*, 61, 273-287.

Lord, F.M. (1952). A Theory of Test Scores. *Psychometric Monograph*, 7.

Luce, R.D. (1959). *Individual Choice Behaviour*. New York, Wiley.

Luce, R.D., Bush, R.R. & Galanter E. (Eds) (1963). *Handbook of Mathematical Psychology*. (3 volumes). New York, Wiley.

Mislevy, R.J. (1984). Estimating Latent Distributions. *Psychometrika*, 49, 359-381.

Neyman, J. & Scott, E.L. (1948). Consistent Estimates, Based on Partially Consistent Observations. *Econometrika*, 16, 1-32.

Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Copenhagen, Danish Institute for Educational Research.

Sternberg, S.H. (1959). A Path Dependent Linear Model. In: R.R. Bush & W.K. Estes (Eds), *Studies in Mathematical Learning Theory*. Stanford, Stanford University Press, 308-339.

Sternberg, S.H. (1963). Stochastic Learning Theory. In: R.D. Luce, R.R. Bush, & E. Galanter (Eds), *Handbook of Mathematical Psychology*, New York, Wiley, Vol. II, 1-120.

Verhelst, N.D. & Eggen, T.J.H.M. (1989), Psychometrische en Statistische Aspecten van Peilingsonderzoek (in Dutch). (Psychometric and Statistical Aspects of Assessment Research.), Arnhem, Cito, PPON-rapport, nr. 4.

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