First publ. in: Mathematical Models for Social Psychology / ed. by Wilhelm F. Kempf ...
Bern : Huber, 1977, pp. 81-100

Conditional Inference for the Dynamic Test Model

WILHELM F. KEMPF AND PETER HAMPAPA

3.1 The Model

A new type of probabilistic test model was developed by KEMPF (1974a, b) and applied to tests in which a correct answer to an item may produce some kind of transfer or learning effect. The basic concept of the new model is the *conditional item characteristic function* $f_{i \cdot s_{vi}}(\xi_v)$ which enables us to replace the usual assumption of local stochastic independence (cf. LORD & NOVICK, 1968, 360f.) by the much less restrictive concept of local serial dependence: an individual's response to an item is assumed to depend on his (her) prior responses in the same test.

Let ξ_v be a latent trait parameter describing the (initial) ability of an individual v in some reasoning test. Then we define a set of random variables a_{vi} (v = 1, n; i = 1, k) such that $a_{vi} = 1$ if the individual v gives a correct response to the i-th item and $a_{vi} = 0$ if the response is incorrect. The *response vector* $(a_{vi}) = (a_{v1}, \ldots, a_{vk})$ gives a complete description of the individual's responses to a test of k items. If i = 1, ...2, k is the sequence in which the items are responded to, then the distribution of the response vector follows from Eq. 3.1:

$$(3.1) \qquad p\{(a_{vi})\} = \prod_{i=1}^{k} p\{a_{vi} \mid a_{v1}, \ldots, a_{v,\,i-1}\}.$$

Now let s_{vi} be the *partial response vector* $(a_{v1}, \ldots, a_{v,i-1})$. Then the conditional item characteristic function is defined by Eq. 3.2:

$$(3.2) \qquad f_{i \cdot s_{vi}}(\xi_v) = p\{a_{vi} = 1 \mid (a_{v1}, \ldots, a_{v,i-1}) = s_{vi}\}.$$

Eq. (3.3) describes the conditional distribution of the item score variable a_{vi}, given the individual's responses to preceding items:

$$(3.3) \qquad p\{a_{vi} \mid s_{vi}\} = f_{i \cdot s_{vi}}(\xi_v)^{a_{vi}} [1 - f_{i \cdot s_{vi}}(\xi_v)]^{1 - a_{vi}}.$$

With regard to the structural form of $f_{i \cdot s_{vi}}(\xi_v)$ as a function of the latent variable ξ, there are various possibilities. Any cumulative distribution function could be chosen. There are some functions, however, which lead to far more attractive statistical models than others.

For psychological reasons, the conditional item characteristic functions will contain at least three types of parameters: *individual parameters* describing the individual's ability, *item parameters* describing the difficulty of the items, and *transfer parameters* that describe how an individual's responses are influenced by his answers to the preceding items. The main statistical problem with models like this is how to handle the individual parameters. Psychological tests are typically administered to a large number of individuals, and since each individual v is characterized by a parameter ξ_v we are faced with an extremly large number of parameters. The individual parameters are a nuisance, however, if we want to make statistical inferences about item difficulties or about transfer processes such as learning or reactive inhibition. One would prefer to eliminate these parameters from the statistical analysis.

This can be done by using *conditional inference,* as suggested by RASCH (1961) and ANDERSEN (1970, 1973a). If sufficient statistics exist for the parameters to be eliminated, then the analysis can be based on a conditional distribution, given the values of these statistics. From the definition of sufficiency it follows that this conditional distribution will no longer depend on the undesirable parameters. In contrast to the direct or unconditional maximum-likelihood method, the conditional approach gives consistent results when the number of individuals is large (cf. ANDERSEN, 1970).

ANDERSEN (1973b) has shown that the models developed by RASCH (1960, 1961) are the only ones that permit this kind of conditional inference, given that local stochastic independence applies. The dynamic test model (KEMPF, 1974a, b) is a natural extension of the Rasch model for binary items.

The conditional item characteristic functions of the dynamic model have the form,

$$(3.4) \qquad f_{i \cdot s_{vi}}(\xi_v) = \frac{\xi_v + \psi_{s_{vi}}}{\xi_v + \sigma_i}, \quad \psi_{s_{vi}} < \sigma_i,$$

where ξ_v is an ability parameter, σ_i is an item difficulty parameter, and $\psi_{s_{vi}}$ is a transfer parameter. For a more exhaustive discussion of the role that the parameters play in the model we refer to chapter 1 of this book.

In KEMPF (1974a, b), the conditional item characteristic functions Eq. (3.4) were treated as dependent on the ***number of correct responses*** to the preceding items,

$$(3.5) \qquad r_{vi} = \begin{cases} 0 & \text{for } i = 1 \\ \sum_{j=1}^{i-1} a_{vj} & \text{for } i = 2, 3, \ldots, k \end{cases}$$

but *not* on *which* of the preceding items were answered correctly. According to this assumption,

$$(3.6) \qquad f_{i \cdot s_{vi}}(\xi_v) = f_{i \cdot r_{vi}}(\xi_v)$$

holds for all partial response vectors s_{vi} which are compatible with the partial score r_{vi}, and the model (Eq. 3.4) is reduced to

$$(3.7) \qquad f_{i \cdot r_{vi}}(\xi_v) = \frac{\xi_v + \psi_{r_{vi}}}{\xi_v + \sigma_i}, \text{ where } \psi_r < \sigma_i.$$

However, with n individual parameters $(\xi_1, \ldots, \xi_n)$, k item parameters $(\sigma_1, \ldots, \sigma_k)$, and k transfer parameters $(\psi_0, \ldots, \psi_{k-1})$, the model (3.7) still contains too many free parameters. We note that $(\xi_v + c) + (\psi_r - c) = \xi_v + \psi_r$, and $(\xi_v + c) + (\sigma_i - c) = \xi_v + \sigma_i$. We note further that Eq. 3.4 remains unchanged if both numerator and denominator are multiplied by some positive constant. We conclude that the parameters are measured on interval scales and hence we can set

$$(3.8) \qquad \prod_{i=1}^{k} \sigma_i = 1 \text{ and } \mathrm{MIN}\,(\psi_r) = 0,$$

which results in an appropriate set of parameters for the model.

If there is no transfer at all, we may set $\psi_0 = \psi_1 = \ldots = \psi_{k-1} = 0$ and Eq. 3.7 is reduced to

$$(3.9) \qquad f_{i \cdot r_{vi}}(\xi_v) = \frac{\xi_v}{\xi_v + \sigma_i} = f_i(\xi_v) \text{ for } r_{vi} = 0, 1, \ldots, k-1,$$

which is the (unconditional) item characteristic function of the Rasch model for binary items.

As KEMPF (1974a) has shown, the extended model (Eq. 2.4) has essentially the same mathematical properties as the simple Rasch model (Eq. 2.6). The raw scores $a_{vo} = \Sigma_{i=1}^{k} a_{vi}$ are sufficient statistics for the latent ability parameters ξ_v, and the item and transfer parameters can be estimated by means of the conditional maximum-likelihood method from the conditional likelihood function

$$(3.10) \qquad L = p\{((a_{vi})) \mid (a_{vo})\} = \prod_{v=1}^{N} p\{(a_{vi}) \mid a_{vo}\},$$

where $((a_{vi}))$ denotes the matrix of the responses of N individuals to k items and (a_{vo}) denotes the vector of the raw scores of the individuals. If $a_{vo} = 0$ or $a_{vo} = k$, then $p\{(a_{vi}) \mid a_{vo}\} = 1$ and hence does not contribute to the conditional likelihood (Eq. 3.10). Individuals with a raw score $a_{vo} = 0$ or $a_{vo} = k$ can therefore be deleted from the sample.

3.2 The Conditional Estimation Equations

Let us consider the responses of n individuals with $0 < a_{vo} < k$. Then, the conditional likelihood of the response matrix $((a_{vi}))$ follows from Eq. 3.11:

$$(3.11) \qquad L = \prod_{v=1}^{n} p\{(a_{vi}) \mid a_{vo}\} = \prod_{v=1}^{n} \frac{p\{(a_{vi})\}}{p\{a_{vo}\}},$$

where $p\{(a_{vi})\}$ is obtained from inserting Eqs. 3.6–3.7 into Eqs. 3.1–3.3 and $p\{a_{vo}\}$ is obtained by summation of the probabilities $p\{(a^*_{vi})\}$ of all possible response vectors (a^*_{vi}) which are compatible with the raw score a_{vo} (i.e., $\Sigma_{i=1}^{k} a^*_{vi} = a_{vo}$):

$$(3.12) \qquad L = \prod_{v=1}^{n} \frac{\prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\sum_{(a^*_{vi}) \mid a_{vo}} \prod_{i=1}^{k} (\sigma_i - \psi_{r^*_{vi}})^{1 - a^*_{vi}}},$$

where $r^*_{vi} = \Sigma_{j=1}^{i-1} a^*_{vi}$ for i = 2, 3, ..., k, and $r^*_{vi} = 0$ for i = 1.

Now let n_{ri} be the number of individuals who give an incorrect answer to item number i after $r_{vi} = r$ correct responses to the preceding items $j = 1, 2, ..., i - 1$. We observe that, for $i = 1, 2, ..., k$ and for $r = 0, 1, ..., i - 1$, the expression $(\sigma_i - \psi_r)$ occurs n_{ri} times in the numerator of Eq. 3.4, each time raised to the power $1 - a_{vi} = 1$, and in all other places where it occurs in the numerator of Eq. 3.4 it is raised to the power $1 - a_{vi} = 0$. Furthermore, let N_{k-s} be the number of individuals who give a total of s incorrect responses to the k items so that $a_{vo} = k - s$. Then we see that, for $s = 1, 2, ..., k - 1$, the expression

$$G(k, s) = \sum_{(a^*_{vi}) \mid k - s} \prod_{i=1}^{k} (\sigma_i - \psi^*_{r_{vi}})^{1 - a^*_{vi}} \tag{3.13}$$

occurs N_{k-s} times in the denominator of Eq. 3.12 and hence we can simplify Eq. 3.12 to

$$L = \frac{\prod_{i=1}^{k} \prod_{r=0}^{i-1} (\sigma_i - \psi_r)^{n_{ri}}}{\prod_{s=1}^{k-1} G(k, s)^{N_{k-s}}}. \tag{3.14}$$

From Eq. 3.14, we finally obtain the conditional estimation equations which must be solved under the side condition of $\psi_\beta < \sigma_\alpha$ for all α and β, by setting $\partial \ln(L)/\partial\sigma_\alpha = 0$ for $\alpha = 1, 2, ..., k$ and $\partial \ln(L)/\partial\psi_\beta = 0$ for $\beta = 0, 1, ..., k - 1$:

$$\sum_{r=0}^{\alpha-1} \frac{n_{r\alpha}}{\sigma_\alpha - \psi_r} - \sum_{s=1}^{k-1} N_{k-s} \frac{\partial G(k, s)/\partial\sigma_\alpha}{G(k, s)} = 0 \text{ for } \alpha = 1, ..., k \tag{3.15}$$

and $$\sum_{i=\beta+1}^{k} \frac{n_{\beta i}}{\psi_\beta - \sigma_i} - \sum_{s=1}^{k-1} N_{k-s} \frac{\partial G(k, s)/\partial\psi_\beta}{G(k, s)} = 0 \text{ for } \beta = 0, ..., k - 1.$$

From Eqs. 3.15 we see that we need to study the properties of the G-functions (Eq. 3.13) before we can solve the estimation equations.

From Eq. 3.5 we see that G(k, s) is a sum of elements which are the products of s factors $(\sigma_i - \psi^*_{r_i})$. Now let σ_{i_j} be the item parameter in the j-th factor. Then $a^*_{vi} = 0$ for j − 1 of the preceding items and, hence, $a^*_{vi} = 1$ for $r^*_{vi_j} = i_j - j$ of the items, $i < i_j$. We note further that a particular item parameter can appear only once in a product, and that the products are summed over all possible combinations of item parameters. Hence, i_j cannot be larger than k − s + j and G(k, s) can be written in the form

$$(3.16) \qquad G(k, s) = \sum_{i_1 = 1}^{k-s+1} \sum_{i_2 = i_1 + 1}^{k-s+2} \cdots \sum_{i_s = i_{s-1}+1}^{k} \prod_{j=1}^{s} (\sigma_{i_j} - \psi_{i_j - j}).$$

In Section 3.4 we give the proof of a theorem according to which

$$(3.17) \qquad G(k, s) = \hat{G}(k, s)$$

with $\hat{G}(k, s)$ being defined as

$$(3.18) \qquad \hat{G}(k, s) = \sum_{m=0}^{s} \delta_m(k-s)\,\gamma_{s-m}(k) \cdot (-1)^m$$

where $\delta_m(k - s)$ denotes the sum of all possible products of m transfer parameters from the set $\psi_0, \ldots, \psi_{k-s}$, with repetitions:

$$(3.19) \qquad \delta_m(k-s) = \begin{cases} 1 \text{ for } m = 0 \\ \displaystyle\sum_{j_1=0}^{k-s} \sum_{j_2=j_1}^{k-s} \cdots \sum_{j_m=j_{m-1}}^{k-s} \prod_{t=1}^{m} \psi_{j_t} \\ \text{for } m = 1, 2, \ldots, s. \end{cases}$$

For example:

$$\delta_0(k-s) = 1$$

$$\delta_1(k-s) = \psi_0 + \psi_1 + \psi_2 + \dots\dots\dots + \psi_{k-s}$$

$$\begin{aligned}\delta_2(k-s) = {} & \psi_0^2 + \psi_0\psi_1 + \psi_0\psi_2 + \dots\dots\dots + \psi_0\psi_{k-s} \\ & + \psi_1^2 + \psi_1\psi_2 + \dots\dots\dots + \psi_1\psi_{k-s} \\ & + \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ & \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ & \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ & + \psi_{k-s-1}^2 + \psi_{k-s-1}\psi_{k-s} + \psi_{k-s}^2\end{aligned}$$

etc.

$\gamma_{s-m}(k)$ denotes the sum of all possible products of s − m of the item parameters $\sigma_1, \dots, \sigma_k$, without repetitions:

$$\text{(3.20)} \qquad \gamma_{s-m}(k) = \begin{cases} \displaystyle\sum_{i_1=1}^{k-s+m+1} \sum_{i_2=i_1+1}^{k-s+m+2} \dots \sum_{i_{s-m}=i_{s-m-1}+1}^{k} \prod_{t=1}^{s-m} \sigma_{i_t} & \text{for } m = 0, 1, \dots, s-1 \\ 1 & \text{for } m = s \end{cases}$$

For example:

$$\gamma_0(k) = 1$$

$$\gamma_1(k) = \sigma_1 + \sigma_2 + \sigma_3 + \dots\dots\dots + \sigma_k$$

$$\begin{aligned}\gamma_2(k) = {} & \sigma_1\sigma_2 + \sigma_1\sigma_3 + \sigma_1\sigma_4 + \dots\dots\dots + \sigma_1\sigma_k \\ & + \sigma_2\sigma_3 + \sigma_2\sigma_4 + \dots\dots\dots + \sigma_2\sigma_k \\ & + \sigma_3\sigma_4 + \dots\dots\dots + \sigma_3\sigma_k \\ & + \dots\dots\dots\dots\dots\dots \\ & \dots\dots\dots\dots\dots\dots \\ & \dots\dots\dots\dots\dots\dots \\ & + \sigma_{k-1}\sigma_k\end{aligned}$$

etc.

In the literature, $\gamma_{s-m}(k)$ is known as the elementary symmetric function of order s − m of the parameters $\sigma_1, \dots, \sigma_k$.

From Eqs. 3.17–3.20 we can easily arrive at the first-order partial derivatives of the G-functions. As we shall see, the partial derivative can also be expressed in terms of δ-functions and γ-functions.

From Eq. 3.19 we get:

$$\psi^0_{k+1-s}\delta_m(k-s) = \sum_{j_1=0}^{k-s} \cdots\cdots \sum_{j_m=j_{m-1}}^{k-s} \prod_{t=1}^{m} \psi_{j_t}$$

$$\psi^1_{k+1-s}\delta_{m-1}(k-s) = \sum_{j_1=0}^{k-s} \cdots\cdots \sum_{j_{m-1}=j_{m-2}}^{k-s} \sum_{j_m=k+1-s}^{k+1-a} \prod_{t=1}^{m} \psi_{j_t}$$

$$\vdots$$

$$\psi^{m-1}_{k+1-s}\delta_1(k-s) = \sum_{j_1=0}^{k-s} \sum_{j_2=k+1-s}^{k+1-s} \cdots\cdots \sum_{j_m=j_{m-1}}^{k+1-s} \prod_{t=1}^{m} \psi_{j_t}$$

$$\psi^{m}_{k+1-s}\delta_0(k-s) = \sum_{j_1=k+1-s}^{k+1-s} \sum_{j_2=j_1}^{k+1-s} \cdots\cdots \sum_{j_m=j_{m-1}}^{k+1-s} \prod_{t=1}^{m} \psi_{j_t}$$

From these expressions, we obtain a useful recursive formula for the δ-functions which is basic to any practical computations in connection with the conditional likelihood (Eq. 3.14):

$$\text{(3.21)} \qquad \sum_{\eta=0}^{m} \psi^{m-\eta}_{k+1-s}\delta_\eta(k-s) = \delta_m(k+1-s).$$

Eq. 3.21 expresses each of the δ-functions in terms of δ-functions based on one parameter less. This makes it possible to start with $\delta_m(0)$, for m = 1, ..., k − 1, and to apply Eq. 3.21 repeatedly until we arrive at $\delta_m(k-1)$ after k − 1 repetitions.

We shall now proceed to the partial derivatives of the δ-functions. From Eq. 3.21 we obtain, for r = k − s and m > 0:

$$\partial\delta_m(k-s)/\partial\psi_{k-s} = \sum_{\eta=0}^{m-1} (m-\eta)\, \psi^{m-\eta-1}_{k-s} \cdot \delta_\eta(k-1-s)$$

and since $\Sigma_{\eta=0}^{m-1}(m-\eta)\,x_\eta = \Sigma_{\eta=0}^{m-1}\Sigma_{\iota=0}^{m-1-\eta}x_\iota$:

$$\partial\delta_m(k-s)/\partial\psi_{k-s} = \sum_{\eta=0}^{m-1}\sum_{\iota=0}^{m-1-\eta}\psi_{k-s}^{m-\iota-1}\delta_\iota(k-1-s)$$
$$= \sum_{\eta=0}^{m-1}\psi_{k-s}^{\eta}\sum_{\iota=0}^{m-1-\eta}\psi_{k-s}^{m-\eta-\iota-1}\delta_\iota(k-1-s)$$
$$= \sum_{\eta=0}^{m-1}\psi_{k-s}^{\eta}\delta_{m-\eta-1}(k-s).$$

Since $\delta_m(k-s)$ is symmetric in its arguments, the result can be generalized and

$$(3.22)\qquad \partial\delta_m(k-s)/\partial\psi_r = \sum_{\eta=0}^{m-1}\psi_r^{\eta}\delta_{m-\eta-1}(k-s)$$

holds for all $r = 0, ..., k-s$ and $m > 0$. For $m = 0$ and for $r > k-s$ the δ-functions do not depend on ψ_r, and $\partial\delta_m(k-s)/\partial\psi_r = 0$.

For the partial derivatives of the G-functions we thus obtain

$$(3.23)\qquad \partial G(k-s)/\partial\psi_r = \begin{cases} \sum_{m=1}^{s}\gamma_{s-m}(k)\left(\sum_{j=0}^{m-1}\psi_r^j\cdot\delta_{m-1-j}(k-s)\right)(-1)^m & \text{for } r = 0, ..., k-s \\ 0 & \text{for } r > k-s \end{cases}$$

and since $\partial\gamma_{s-m}(k)/\partial\sigma_i = \gamma_{s-m-1}^{(i)}(k)$ for $m < s$ and $\partial\gamma_{s-m}(k)/\partial\sigma_i = 0$ for $m = s$,

$$(3.24)\qquad \partial G(k-s)/\partial\sigma_i = \sum_{m=0}^{s-1}\delta_m(k-s)\,\gamma_{s-m-1}^{(i)}(k)\cdot(-1)^m$$

where $\gamma_{s-m-1}^{(i)}(k)$ denotes the elementary symmetric function of order $s-m-1$ of the parameters $\sigma_1, ..., \sigma_{i-1}, \sigma_{i+1}, ..., \sigma_k$. Finally, the conditional estimation equations assume the form, (3.25) $\partial \ln(L)/\partial\sigma_\alpha =$

$$\sum_{r=0}^{\alpha-1} \frac{n_{r_\alpha}}{\sigma_\alpha - \psi_r} - \sum_{s=1}^{k-1} N_{k-s} \frac{\sum_{m=0}^{s-1} \delta_m(k-s)\,\gamma^{(\alpha)}_{s-m-1}(k)\,(-1)^m}{\sum_{m=0}^{s} \delta_m(k-s)\,\gamma_{s-m}(k)\,(-1)^m} = 0$$

for $\alpha = 1, ..., k$ and

$\partial \ln(L)/\partial\psi_\beta =$

$$\sum_{i=\beta+1}^{k} \frac{n_{\beta i}}{\psi_\beta - \sigma_i} - \sum_{\substack{s=1 \\ s \leqslant k-\beta}}^{k-1} N_{k-s} \frac{\sum_{m=1}^{s} \gamma_{s-m}(k)\,(\sum_{j=0}^{m-1} \psi_\beta^j \delta_{m-1-j}(k-s))\cdot(-1)^m}{\sum_{m=0}^{s} \delta_m(k-s)\,\gamma_{s-m}(k)\cdot(-1)^m} = 0$$

for $\beta = 0, ..., c_{max}$,

where c_{max} denotes the largest observed raw score $a_{vo} < k$.

The remaining problem connected with the solution of Eq. 3.25 is how to handle the side conditions $\psi_r < \sigma_i (r - 0, ..., k-1; i = 1, ..., k)$ in a proper way. One possibility is to apply joint linear parameter transformations $\psi_r \rightarrow \psi_r^*$ and $\sigma_i \rightarrow \sigma_i^*$, so that $0 < \psi_r^* < 1 \leqslant \sigma_i^*$ $(r = 0, ..., k-1, i = 1, ..., k)$, and to introduce auxiliary parameters $\Phi_r = \ln(\psi_r^*/(1 - \psi_r^*))$ and $\eta_i = \ln(\sigma_i^* - 1)$. The solutions to Eq. 3.25 can then be computed by

(3.26) $\psi_r^* = \exp(\Phi_r)/(1 + \exp(\Phi_r))$

and

(3.27) $\sigma_i^* = 1 = \exp(\eta_i)$

from the solutions to $\partial\ln(L)/\partial\eta_\alpha = (\partial\ln(L)/\partial\sigma_\alpha)(\partial\sigma_\alpha/\partial\eta_\alpha) = 0$ and $\partial\ln(L)/\partial\Phi_\beta = (\partial\ln(L)/\partial\psi_\beta)(\partial\psi_\beta/\partial\Phi_\beta) = 0$.

As a convenient side condition for the auxiliary parameters we set

(3.28) $\text{MIN}(\psi_r^*) = 1 - \text{MAX}(\psi_r^*) = \text{MIN}(\sigma_i^*) - 1.$

A FORTRAN-program for the numerical computation of the parameter estimates is available from Kempf & Mach (1975). Starting with the initial values $\psi_0^* = \psi_1^* = ... = \psi_{k-1}^* = 0.5$ and $\sigma_1^* = \sigma_2^* = ... = \sigma_k^* = 1.5$, the program applies a gradient method (Fischer & Formann, 1972) to compute an iterative solution to the equation system.

The following example serves to illustrate the procedure. The data stem from a survey on sex role differentiation and were taken from DUNCAN (1973). 603 female persons, mothers of children under 19 years old, were asked which of the following activities should be performed regularly by a boy, by a girl, or by both.

1. Shoveling walks.
2. Washing the car.
3. Dusting furniture.
4. Making beds.

The responses were scored $a_{vi} = 1$ if the answer was "both" and $a_{vi} = 0$ if the answer was "boy" or "girl". A large raw score a_{vo} thus expresses low sex role discrimination and the hypothesis is that the answers will produce positive transfer so that an examinee who already has expressed equality of sexes with respect to a majority of the items will show an even stronger tendency to do so with the remaining items.

The original data are summarized in Table 3.1.

After omitting the responses of all examinees with $a_{vo} = 0$ or $a_{vo} = k$, we compute the raw score frequencies N_{k-s}, the item marginals $a_{oi} = \Sigma_{v=1}^{n} a_{vi}$ and the corresponding $((n_{ri}))$ matrix for the remaining 359 examinees (Table 3.2).

Finally, Table 3.3 shows the conditional maximum-likelihood (CML) estimates for the data in Table 3.2, obtained from the computer program by KEMPF & MACH (1975).

Table 3.1: Four-way tabulation of the Duncan (1973) data

		Dust Beds	0. Girl 0. Girl	0. Girl 1. Both	1. Both 0. Girl	1. Both 1. Both
Walks	*Car*					
0. Boy	0. Boy		82	49	1	18
0. Boy	1. Both		40	67	2	38
1. Both	0. Boy		10	12	0	6
1. Both	1. Both		32	80	4	153

Other answers: 9[1]

[1] Includes "girl" for Walks or Car, "boy" for Dust or Beds, "neither" for any item or no answer for one or more items.

Table 3.2: Raw score frequencies, item marginals, and the $((n_{ri}))$-matrix for the Duncan data

s	1	2	3	
N_{k-s}	128	131	100	
i	1	2	3	4
a_{oi}	144	263	69	270
r = 0	215	68	49	0
r = 1		28	129	51
r = 2	$n_{ri} =$		112	34
r = 3				4

Table 3.3: CML estimates for the Duncan data

r	$\hat{\psi}_r$	$\hat{\psi}^*_r$	$\hat{\Phi}_r$	i	$\hat{\sigma}_i$	$\hat{\sigma}^*_i$	$\hat{\eta}_i$
0	0.0000	0.0021	−6.1613	1	1.0564	2.2574	0.2290
1	0.2606	0.5584	0.2348	2	0.4684	1.0021	−6.1634
2	0.2708	0.5801	0.3231	3	3.2280	6.8933	1.7738
3	0.4665	0.9979	6.1613	4	0.6260	1.3386	−1.0829

3.4 Proof of Theorem 3.17[1]

1. As the first step of the proof we show that

   $G(k, s) = \hat{G}(k, s)$ holds for $s = 1$.

   From Eq. 3.16 we obtain

   $$(3.29)\quad G(k, 1) = \sum_{i=1}^{k} (\sigma_i - \psi_{i-1})$$

   and since $\delta_0(k-1) = 1$, $\delta_1(k-1) = \sum_{r=0}^{k-1} \psi_r$, $\gamma_0(k) = 1$ and $\gamma_1(k) =$
$= \sum_{i=1}^{k} \sigma_i$, we may write

[1] Reading of this section can be omitted without loss of continuity.

$$(3.30)\quad G(k, 1) = \delta_0(k-1)\,\gamma_1(k) - \delta_1(k-1)\cdot\gamma_0(k) = \hat{G}(k, 1).$$

This completes the first step of the proof.

2. As the second step of the proof we show that

   $G(k, k) = \hat{G}(k, k)$ holds for all $k \epsilon N$.

   From Eq. 3.30 $G(1, 1) = \hat{G}(1, 1)$. Assume that $G(k, k) = \hat{G}(k, k)$. Then

$$(3.31)\quad G(k, k) = \prod_{i=1}^{k} (\sigma_i - \psi_0)$$

and

$$(3.32)\quad G((k+1), (k+1)) = \prod_{i=1}^{k+1} (\sigma_i - \psi_0) = G(k, k)\,(\sigma_{k+1} - \psi_0).$$

Finally, since $\delta_m(0) = \psi_0^m$

$$(3.33)\quad \hat{G}(k, k)\,(\sigma_{k+1} - \psi_0) = \sum_{m=0}^{k} (-\psi_0)^m \gamma_{k-m}(k)\,(\sigma_{k+1} - \psi_0)$$

$$= \sum_{m=0}^{k} (-\psi_0)^m \gamma_{k-m}(k)\,\sigma_{k+1} + \sum_{m=0}^{k} (-\psi_0)^{m+1} \gamma_{k-m}(k)$$

$$= \gamma_k(k)\,\sigma_{k+1} + \sum_{m=1}^{k} (-\psi_0)^m \gamma_{k-m}(k)\,\sigma_{k+1} + \sum_{m=1}^{k+1} (-\psi_0)^m \gamma_{k+1-m}(k)$$

$$= \gamma_k(k)\,\sigma_{k+1} + \sum_{m=1}^{k} (-\psi_0)^m \gamma_{k-m}(k)\,\sigma_{k+1} + \sum_{m=1}^{k} (-\psi_0)^m \gamma_{k+1-m}(k) + (-\psi_0)^{k+1}$$

$$= \gamma_{k+1}(k+1) + \sum_{m=1}^{k} ((-\psi_0)^m (\gamma_{k-m}(k)\,\sigma_{k+1} + \gamma_{k+1-m}(k))) + (-\psi_0)^{k+1}$$

$$= \gamma_{k+1}(k+1) + \sum_{m=1}^{k} (-\psi_0)^m \gamma_{k+1-m}(k+1) + (-\psi_0)^{k+1}$$

$$= \sum_{m=0}^{k+1} (-\psi_0)^m \gamma_{k+1-m}(k+1)$$

$$= \hat{G}((k+1), (k+1))$$

and hence

(3.34) $\hat{G}((k+1),(k+1)) = G((k+1),(k+1))$,

which completes the second step of the proof.

3. For the last step of the proof we make use of the recursion formula,

(3.35) $G((k+1),s) = G(k,s) + (\sigma_{k+1} - \psi_{k+1-s})\, G(k,(s-1))$

which follows immediately from

(3.36) $$G(k,s) \sum_{i_1=1}^{k-s+1} \cdots \sum_{i_s=i_{s-1}+1}^{k} \prod_{j=1}^{s} (\sigma_{i_j} - \psi_{i_j - j})$$

and

(3.37) $(\sigma_{k+1} - \psi_{k+1-s})\, G(k,(s-1)) =$

$$= (\sigma_{k+1} - \psi_{k+1-s}) \sum_{i_1=1}^{k-s+2} \cdots \sum_{i_{s-1}=i_{s-2}}^{k} \prod_{j=1}^{s-1} (\sigma_{i_j} - \psi_{i_j - j})$$

and we show that the same recursion formula holds for $\hat{G}((k+1),s)$.
Eq. 3.18 yields

(3.38) $$\hat{G}((k+1),s) = \sum_{m=0}^{s} \delta_m(k+1-s)\, \gamma_{s-m}(k+1)\, (-1)^m$$

and since $\gamma_{s-m}(k+1) = \gamma_{s-m}(k) + \gamma_{s-m-1}(k)\, \sigma_{k+1}$

$$= \sum_{m=0}^{s} \delta_m(k+1-s)\, \gamma_{s-m}(k)\, (-1)^m + \sum_{m=0}^{s} \delta_m(k+1-s) \gamma_{s-m-1}(k) \sigma_{k+1} (-1)^m$$

$$= \sigma_{k+1} \hat{G}(k,(s-1)) + \sum_{m=0}^{s} \delta_m(k+1-s)\, \gamma_{s-m}(k)\, (-1)^m.$$

From the recursion formula Eq. 3.21 we obtain

(3.39) $$\delta_m(k+1-s) - \delta_m(k-s) = \sum_{\eta=0}^{m-1} \psi_{k+1-s}^{m-\eta}\, \delta_\eta(k-s)$$

and

$$(3.40)\quad \delta_{m-1}(k+1-s) = \sum_{\eta=0}^{m-1} \psi_{k+1-s}^{m-\eta-1}\,\delta_{\eta}(k-s)$$

and hence

$$(3.41)\quad \delta_{m}(k+1-s) = \delta_{m}(k-s) + \psi_{k+1-s}\,\delta_{m-1}(k+1-s).$$

With the definition $\delta_{-1}(k+1-s) = 0$, we may therefore write

$$(3.42)\quad \hat{G}((k+1), s) =$$

$$= \sigma_{k+1}\cdot\hat{G}(k,(s-1)) + \sum_{m=0}^{s} (\delta_{m}(k-s) + \psi_{k+1-s}\,\delta_{m-1}(k+1-s))\,\gamma_{s-m}(k)\cdot(-1)^{m}$$

$$= \sigma_{k+1}\,\hat{G}(k,(s-1)) + \sum_{m=0}^{s} \delta_{m}(k-s)\,\gamma_{s-m}(k)\,(-1)^{m} +$$

$$+ \psi_{k+1-s} \sum_{m=0}^{s} \delta_{m-1}(k+1-s)\,\gamma_{s-m}(k)\,(-1)^{m}$$

$$= \sigma_{k+1}\cdot\hat{G}(k,(s-1)) + \hat{G}(k,s) + \psi_{k+1-s} \sum_{m=1}^{s} \delta_{m-1}(k+1-s)\,\gamma_{s-m}(k)\,(-1)^{m}$$

$$= \sigma_{k+1}\cdot\hat{G}(k,(s-1)) + \hat{G}(k,s) + \psi_{k+1-s} \sum_{m=0}^{s-1} \delta_{m}(k+1-s)\,\gamma_{s-m-1}(k)\,(-1)^{m-1}$$

$$= \sigma_{k+1}\cdot\hat{G}(k,(s-1)) + \hat{G}(k,s) - \psi_{k+1-s}\,\hat{G}(k,(s-1)),$$

which completes the proof of the theorem.

3.5 A Conditional Goodness-of-Fit Test

An important step in any statistical analysis is the determination of how well the model fits the data. For the present model, this step can be carried out in form of a likelihood-ratio test based on an approximation to the χ^2-distribution. The rationale for the goodness-of-fit test derives from the conditional approach.

$$(3.43) \qquad p\{(a_{vi}) \mid a_{vo} = k - s\} = \frac{\prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\sum_{m=0}^{s} \delta_m(k - s)\, \gamma_{s-m}(k) \cdot (-1)^m}.$$

Eq. 3.43 shows that the distribution of an examinee's responses, given his or her raw score a_{vo}, is independent of the latent trait parameter ξ_v and depends on the item and transfer parameters only. The latter can be estimated from any subgroup G_ν of examinees by taking the product of Eq. 3.43 over the examinees of this subgroup as our conditional likelihood $L_\nu = \prod_{v \epsilon G_\nu} p\{(a_{vi}) \mid (a_{vo}\}$.

If $G_1, ..., G_M$ are M disjoint subgroups of examinees, we define restricted CML estimates of the auxiliary parameters as the solutions to the restricted likelihood equations,

$$(3.44) \qquad \partial \ln(L_\nu) / \partial \Phi_r = 0 \qquad \text{for } r = 0, ..., c_\nu$$

$$\text{and} \qquad \partial \ln(L_\nu) / \partial \eta_i = 0 \qquad \text{for } i = 1, ..., k$$

where c_ν is the largest raw score $a_{vo} < k$ observed in subgroup G_ν. The solutions of Eq. 3.44 are $\hat{\Phi}_0^{(\nu)}, ..., \hat{\Phi}_c^{(\nu)}$ and $\hat{\eta}_1^{(\nu)}, ..., \hat{\eta}_k^{(\nu)}$.

If the model holds, we should always estimate the same parameters $\Phi_r^{(\nu)} = \Phi_r$ and $\eta_i^{(\nu)} = \eta_i$, regardless of wich subset of examinees is selected. This offers the possibility of checking the fit of the model through a comparison of the restricted CML estimates $(\hat{\Phi}_0^{(\nu)}, ..., \hat{\Phi}_c^{(\nu)}; \hat{\eta}_1^{(\nu)}, ..., \hat{\eta}_k^{(\nu)})$, $\nu = 1, ..., M$, with the unrestricted CML estimates $(\hat{\Phi}_0, ..., \hat{\Phi}_{c_{max}};$ $\hat{\eta}_1, ..., \hat{\eta}_k)$.

Let $L(\Phi_0, ..., \Phi_{c_{max}}; \eta_1, ..., \eta_k)$ be the conditional likelihood $\prod_{v=1}^{n} p\{(a_{vi}) \mid a_{vo}\}$ and let $L^{(\nu)}(\Phi_0, ..., \Phi_c; \eta_1, ..., \eta_k)$ be the corresponding restricted conditional likelihood $\prod_{v \epsilon G_\nu} p\{(a_{vi}) \mid a_{vo}\}$. Then we define the conditional likelihood ratio λ by

$$(3.45) \qquad \lambda = \frac{L(\hat{\Phi}_0, \ldots, \hat{\Phi}_{c_{max}}; \hat{\eta}_1, \ldots, \hat{\eta}_k)}{\prod_{\nu=1}^{M} L^{(\nu)}(\hat{\Phi}_0^{(\nu)}, \ldots, \hat{\Phi}_{c_\nu}^{(\nu)}; \hat{\eta}_1^{(\nu)}, \ldots, \hat{\eta}_k^{(\nu)})} .$$

Since $L(\Phi_0, \ldots, \Phi_{c_{max}}; \eta_1, \ldots, \eta_k) = \prod_{\nu=1}^{M} L^{(\nu)}(\Phi_0, \ldots, \Phi_{c_\nu}; \eta_1, \ldots, \eta_k)$ and since $(\hat{\Phi}_0^{(\nu)}, \ldots, \hat{\Phi}_{c_\nu}^{(\nu)}; \hat{\eta}_1^{(\nu)}, \ldots, \hat{\eta}_1^{(\nu)})$ maximizes the ν-th factor in the denominator of Eq. 3.45 we find that $\lambda \leqslant 1$. If the model holds, then the restricted CML estimates should differ only slightly from the overall estimates. Values of λ close to 1 therefore indicate a good fit of the model, and if λ is substantially smaller than 1 the model will be rejected. From a theorem by ANDERSEN (1971), it follows that the distribution of $-2\ln(\lambda)$ converges for $n \to \infty$ to a χ^2-distribution with

$$df = (k-1)(M-1) + \sum_{\nu=1}^{M} c_\nu - c_{max}$$

degrees of freedom. The model will be rejected at an asymptotic significance level α when $-2\ln(\lambda)$ is larger than the $(1-\alpha)$ percentile of the limiting χ^2-distribution. As usual, the degrees of freedom are the number of free parameters specified under the hypothesis.

As an illustration of the technique, we group the examinees of the Duncan survey according to the amount of sex role differentiation expressed in their responses. Group 1 is formed of examinees with high or medium sex role differentiation $1 \leqslant a_{vo} \leqslant 2$. Group 2 contains examinees who show low sex role differentiation $a_{vo} = 3$. The restricted CML estimates for the groups are summarized in Table 3.4.

Table 3.4: Restricted CML estimates for the Duncan data

r	$\hat{\Phi}_r^{(1)}$	$\hat{\Phi}_r^{(2)}$	i	$\hat{\eta}_i^{(1)}$	$\hat{\eta}_i^{(2)}$
0	−5.2096	−0.4210	1	−2.1108	1.1337
1	1.1992	0.9873	2	−5.2150	−1.1567
2	5.2096	−1.4160	3	1.9363	1.9470
3		1.4160	4	−0.6915	−1.6332

Table 3.5: Computation of $-2\ln(\lambda)$ for the Duncan data

$\ln(L^{(v)}(\hat{\Phi}_0^{(\nu)}, \ldots, \hat{\Phi}_{c_\nu}^{(\nu)}; \hat{\eta}_1^{(\nu)}, \ldots, \hat{\eta}_k^{(\nu)}))$	
v = 1	−270.5633
v = 2	−115.9737
$\ln(L(\hat{\Phi}_0, \ldots, \hat{\Phi}_{c_{max}}; \hat{\eta}_1, \ldots, \hat{\eta}_k))$	
	−391.3204
$-2\ln(\lambda)$ =	9.5667

The corresponding test statistic $-2\ln(\lambda)$ is computed in Table 3.5.

With df = (3) (1) + (3 + 2) − 3 = 5 degrees of freedom, the test statistic is between the 90% and the 95% percentiles of the limiting χ^2-distribution. The model is thus on the edge of being rejected.

3.6 Reduction to the Rasch Model

From an inspection of the parameter estimates in Table 3.3, we note that the variability of the transfer parameters ψ_r is rather small relative to the variability of the item difficulties σ_i. Hence, we may ask whether the data allow for a reduction in the number of parameters, assuming

$$(3.46) \qquad \psi_0 = \psi_1 = \ldots = \psi_{c_{max}} = 0.$$

This hypothesis can be tested by a conditional likelihood-ratio test.

Under the null hypothesis (Eq. 3.46), the conditional likelihood (Eq. 3.14) is reduced to

$$(3.47) \qquad L_0(\sigma_1, \ldots, \sigma_k \mid \psi_0 = \ldots = \psi_{c_{max}} = 0) = \frac{\prod_{i=1}^{k} \sigma_i^{n - a_{oi}}}{\prod_{s=1}^{k-1} \gamma_s(k)^{N_{k-s}}}$$

and we obtain CML estimates for the σ's by maximizing Eq. 3.47. By logarithmic differentiation with respect to the σ's we get the conditional estimation equations,

$$\frac{n - a_{0\alpha}}{\sigma_\alpha} - \sum_{s=1}^{k-1} N_{k-s} \frac{\gamma_{s-1}^{(\alpha)}(k)}{\gamma_s(k)} \qquad \text{for } \alpha = 1, ..., k. \tag{3.48}$$

i	$\hat{\sigma}_i^{(0)}$
1	1.6431
2	0.4071
3	4.0241
4	0.3715

Table 3.6: CML estimates of item parameters for the Duncan data under the null hypothesis $\psi_0 = \psi_1 = \psi_2 = \psi_3 = 0$

Table 3.6 shows the CML estimates $\hat{\sigma}_1^{(0)}, ..., \hat{\sigma}_k^{(0)}$ obtained as the solutions to Eq. 3.48 for the Duncan data.

We may then consider the conditional likelihood ratio

$$\lambda^* = \frac{L_0(\hat{\sigma}_1^{(0)}, ..., \hat{\sigma}_k^{(0)} \mid \psi_0 = ... = \psi_{c_{max}} = 0)}{L(\hat{\Phi}_0, ..., \hat{\Phi}_{c_{max}}; \hat{\eta}_1, ..., \hat{\eta}_k)} \tag{3.49}$$

for a comparison of the models in Eqs. 3.7 and 3.9, respectively. The null hypothesis will be rejected when λ^* is substantially smaller than one. Again, we make use of the theorem of ANDERSEN (1971) and approximate the distribution of $2\ln(\lambda^*)$ by a χ^2-distribution with df = c_{max} degrees of freedom. Hence, the null hypothesis (3.46) can be rejected at an asymptotic significance level α when $-2\ln(\lambda^*)$ is larger than the $(1 - \alpha)$ percentile of the χ^2-distribution with df = c_{max} degrees of freedom.

An actual computation of $-2\ln(\lambda^*)$ for the Duncan data gives the result $-2\ln(\lambda^*) = 13.063$ which lies beyond the 99% percentile of the χ^2-distribution with df = 3. The null hypothesis Eq. 3.46 can thus be rejected with asymptotic significance level 1%, and we conclude that the transfer effects observed in the Duncan data must not be neglected. In this case, the dynamic test model Eq. 3.7 cannot be reduced to the Rasch model Eq. 3.9.

References

ANDERSEN, E. B.: Asymptotic properties of conditional maximum likelihood estimates. Journal of the Royal Statistical Society B *32*, 283–301, 1970.

– The asymptotic distribution of conditional maximum likelihood ratio tests. Journal of the American Statistical Association *66*, 630–633, 1971.

– Conditional Inference and Models for Measuring. Mentalhygiejnisk Forlag, Copenhagen 1973a.

– Conditional inference for multiple choice questionnaires. British Journal of Mathematical and Statistical Psychology *26*, 31–44, 1973b.

DUNCAN, O. D.: Preliminary analysis of data on sex role socialization. Unpublished report to the Russell Sage Foundation, 1973.

FISCHER, G. H., FORMANN, A. K.: An algorithm and a FORTRAN program for estimating the item parameters of the linear logistic test model. Research Bulletin No. 11. Department of Psychology, University of Vienna 1972.

KEMPF, W. F.: Dynamische Modelle zur Messung sozialer Verhaltensdispositionen. In: Kempf, W. F. (ed.): Probabilistische Modelle in der Sozialpsychologie. Huber, Bern/Stuttgart/Wien 1974a.

– A dynamic model for measuring individual and situative influences on social behavior. In Compstat 1974. Proceedings in Computational Statistics. Physica-Verlag, Vienna 1974b.

KEMPF, W. F., MACH, G.: A FORTRAN program for CML-estimation in a dynamic test model. In: Kempf, W. F., Hampapa, P., Mach, G. (eds.): Conditional Maximum Likelihood Estimation for a Dynamic Test Model. IPN-Arbeitsbericht. Institut für die Pädagogik der Naturwissenschaften, Kiel 1975.

LORD, F. M., NOVICK, M. R.: Statistical Theories of Mental Test Scores. Addison-Wesley, Reading, Mass. 1968.

RASCH, G.: Probabilistic Models for Some Intelligence and Attainment Tests. Nielsen & Lydiche, Copenhagen 1960.

– On general laws and the meaning of measurement in psychology. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Vol. IV. University of California Press, Berkeley 1961.