The Numerical Solution of a Set of Conditional Estimation Equations

By E. B. ANDERSEN

Copenhagen School of Economics and Business Administration

[Received November 1969. Final revision September 1971]

SUMMARY

In this paper we discuss the numerical solution of a set of conditional likelihood equations essential for the statistical analysis of psychological questionnaires. The model underlying this type of statistical analysis is described in detail and the conditional maximum-likelihood method that is used for estimating item parameters is discussed in relation to the chosen model. The conditional likelihood equations contain a generalized version of the elementary symmetric functions. The main problem in solving the estimation equations consists in finding a rapid recursive way to compute these functions. Such a procedure is described in the paper. In the final section of the paper the developed method is illustrated by a numerical example.

Keywords: COMPUTATION MAXIMUM-LIKELIHOOD ESTIMATES; RASCH MODEL; SOLUTION ESTIMATION EQUATIONS

1. INTRODUCTION

A MODEL for analysing psychological questionnaires was proposed by Rasch (1961). A simple estimation method for the parameters of the model was given by Andersen (1964). In this paper, the exact maximum-likelihood estimates are derived and a complete program for computing the estimates is outlined. In Section 2 the main features of the model is reviewed. In Section 3 the estimation equations are derived. The problem of solving the equations is essentially a problem of an economical way to compute an extended version of the elementary symmetrical function. This problem is treated in Section 4 with a numerical illustration. Section 5 contains an example of the application of the method to data from a questionnaire with 4 questions and 3 answer categories.

2. THE MODEL

We consider the following experiment. N persons answer a questionnaire consisting of k different items. The answer to each item must fall in exactly one of m possible answer categories. The model, as formulated by Rasch (1961), takes into account both item parameters and person parameters such that the probabilities with which a specific person on a specific item chooses between the answer categories depend on a person parameter θ and an item parameter ϵ. We thus have N person parameters $\theta_1, \ldots, \theta_N$ and k item parameters $\epsilon_1, \ldots, \epsilon_k$. Suppose now that the answer categories represent m various ways to express positive and negative reactions to the items. We can, for instance, let the choice of answer-category 1 and m represent, respectively, the most positive and negative reactions to an item and let answer category 2 to $m-1$ represent intermediate reactions. We shall assume that both θ

and ϵ are real-valued parameters. Although many other interpretations are possible in different contexts we shall for convenience suppose that the person parameter is a measure of how positive an attitude a given person has towards the subject of the questionnaire. Thus persons with high values of θ are likely to give positive reactions to the items. All items do not, on the other hand, provoke positive reactions equally often. We shall assume that the item parameters measure the tendency of the different items to provoke positive reactions. A main characteristic of the Rasch model is now that it is invariant towards certain simultaneous transformations of the θ's and ϵ's. We can, for example, expect the same reactions from a very positive person on an item with low tendency to provoke positive reactions as from a very negative person on an item with very high tendency to provoke positive reactions. In fact it is natural to assume that for each combination (θ, ϵ) and a given $\epsilon' \neq \epsilon$, it is possible to find a value θ' such that we can expect the same reaction from a person with parameter θ on an item with parameter ϵ as from a person with parameter θ' on an item with parameter ϵ'. This will certainly be the case if by a proper parametrization we let the probability distribution of the reactions be a function of θ and ϵ, only through $\theta+\epsilon$. In Rasch (1960, Chapter VII) a more formal development and detailed discussion of this property of the model is given.

Let $p_\nu(\theta, \epsilon)$ be the probability that a person with parameter θ uses answer category ν on an item with parameter ϵ. The model is now given by

$$p_\nu(\theta, \epsilon) = \{c(\theta+\epsilon)\}^{-1} \exp\{\phi_\nu(\theta+\epsilon)+\rho_\nu\}, \tag{1}$$

where

$$c(\theta+\epsilon) = \sum_{\mu=1}^{m} \exp\{\phi_\mu(\theta+\epsilon)+\rho_\mu\}. \tag{2}$$

In this model the answer categories are associated with two sets of parameters $\phi_1, \ldots, \phi_m$ and $\rho_1, \ldots, \rho_m$. To illustrate the role of the category parameters, we consider a reference item with parameter $\epsilon = 0$. If we further put $\theta = 0$ in (1) we get the m basic probabilities

$$p_{\nu 0} = p_\nu(0,0) = \exp(\rho_\nu) \Big/ \sum_{\mu=1}^{m} \exp(\rho_\mu), \quad \nu = 1, \ldots, m.$$

Thus $\rho_1, \ldots, \rho_m$ gives the order of magnitude of the m probabilities for a neutral person on a neutral item.

For lack of a better name we shall simply call the ρ's "category parameters".

For $\epsilon = 0$ and $\theta \neq 0$, (1) takes the form

$$q_\nu(\theta) = p_\nu(\theta, 0) = \exp(\phi_\nu\, \theta + \rho_\nu) \Big/ \sum_{\mu=1}^{m} \exp(\phi_\mu\, \theta + \rho_\mu),$$

which shows that $q_\nu(\theta) \to 1$ as $\theta \to \infty$, when ϕ_ν is the largest ϕ and that $q_\mu(\theta) \to 1$ as $\theta \to -\infty$, when ϕ_μ is the smallest ϕ. Thus the largest and smallest ϕ correspond to the answer categories that the extremely positive and extremely negative persons will choose almost certainly. Similar arguments for the intermediate ϕ's show that the ϕ's represent a scaling of the answer categories on a positive–negative scale for the reactions. In psychometric terminology such a scaling of answer categories is called "a scoring". Accordingly we shall term the ϕ's "scoring parameters".

In (1) we can obviously put $\phi_m = 0$ and $\rho_m = 0$, without loss of generality so that

$$p_m(\theta, \epsilon) = \{c(\theta+\epsilon)\}^{-1} \tag{3}$$

and

$$c(\theta+\epsilon) = 1 + \sum_{j=1}^{m-1} \exp\{\phi_j(\theta+\epsilon)+\rho_j\}. \tag{4}$$

For convenience we shall adopt the following notational system. Let $X_{ij}^{(\nu)}$ be a random variable with possible values 1 and 0. Let $x_{ij}^{(\nu)}$ be the corresponding observed value, where

$$x_{ij}^{(\nu)} = \begin{cases} 1 & \text{when person } i \text{ answers question } j \text{ by using category } \nu, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

It then follows immediately from (1) and (2) that

$$P\{(X_{ij}^{(1)}, \ldots, X_{ij}^{(m)}) = (x_{ij}^{(1)}, \ldots, x_{ij}^{(m)})\} = c^{-1}(\theta_i, \epsilon_j) \exp\left[\sum_{\nu=1}^{m} \{\phi_\nu(\theta_i+\epsilon_j)+\rho_\nu\} x_{ij}^{(\nu)}\right]. \tag{6}$$

If we write $\mathbf{X}_{ij} = (X_{ij}^{(1)}, \ldots, X_{ij}^{(m)})$, $\mathbf{x}_{ij} = (x_{ij}^{(1)}, \ldots, x_{ij}^{(m)})$ and use the notations $\{\mathbf{X}_{ij}\}$ and $\{\mathbf{x}_{ij}\}$ for the $N \times k$ matrices with typical elements $\mathbf{X}_{ij}$ and $\mathbf{x}_{ij}$, it follows from (6) that the simultaneous distribution of the observations is given by

$$\begin{aligned} f(\{\mathbf{x}_{ij}\}) &= P(\{\mathbf{X}_{ij}\} = \{\mathbf{x}_{ij}\}) \\ &= \left\{\prod_{i=1}^{N}\prod_{j=1}^{k} c(\theta_i+\epsilon_j)\right\}^{-1} \exp\left[\sum_{i=1}^{N}\sum_{j=1}^{k}\sum_{\nu=1}^{m} \{\phi_\nu(\theta_i+\epsilon_j)+\rho_\nu\} x_{ij}^{(\nu)}\right] \\ &= \left\{\prod_{i=1}^{N}\prod_{j=1}^{k} c^{-1}(\theta_i+\epsilon_j)\right\} \exp\left[\sum_{i=1}^{N}\sum_{\nu=1}^{m} \phi_\nu\,\theta_i\, x_{i.}^{(\nu)} + \sum_{j=1}^{k}\sum_{\nu=1}^{m} \phi_\nu\,\epsilon_j\, x_{.j}^{(\nu)} + \sum_{\nu=1}^{m} \rho_\nu\, x_{..}^{(\nu)}\right], \end{aligned} \tag{7}$$

where $x_{i.}^{(\nu)} = \sum_{j=1}^{k} x_{ij}^{(\nu)}$ is the total number of times person i has used answer category ν, $x_{.j}^{(\nu)}$ is the total number of times the νth category on question j has been used by the N persons, and $x_{..}^{(\nu)}$ is the total number of times category ν has been used. For the marginals we adopt the notation

$$(\mathbf{X}_{i.}) = (\mathbf{X}_{1.}, \ldots, \mathbf{X}_{N.}),$$

where

$$\mathbf{X}_{i.} = (\mathbf{X}_{i.}^{(1)}, \ldots, \mathbf{X}_{i.}^{(m)}),$$

$$(\mathbf{X}_{.j}) = (\mathbf{X}_{.1}, \ldots, \mathbf{X}_{.k}),$$

where

$$\mathbf{X}_{.j} = (\mathbf{X}_{.j}^{(1)}, \ldots, \mathbf{X}_{.j}^{(m)}),$$

and similar expressions for $(\mathbf{x}_{i.})$ and $(\mathbf{x}_{.j})$.

The problem we are interested in is the estimation of the ϵ's and the ϕ's. We shall start, however, by considering a broader class of models. If we put $\theta_{i\nu} = \theta_i\phi_\nu + \rho_\nu$ and $\epsilon_{j\nu} = \epsilon_j\phi_\nu$ in (7) we get

$$f(\{\mathbf{x}_{ij}\}) = \left\{\prod_{i=1}^{N}\prod_{j=1}^{k} c_{ij}^{-1}\right\} \exp\left(\sum_{i=1}^{N}\sum_{\nu=1}^{m} \theta_{i\nu}\, x_{i.}^{(\nu)} + \sum_{j=1}^{k}\sum_{\nu=1}^{m} \epsilon_{j\nu}\, x_{.j}^{(\nu)}\right), \tag{8}$$

where $c_{ij} = c(\theta_i + \epsilon_j)$ is given by (2). Corresponding to $\phi_m = 0$ and $\rho_m = 0$ we have $\theta_{im} = \epsilon_{jm} = 0$ for all j and i. From (7) it follows further that we can introduce the condition $\sum_{j=1}^{k} \epsilon_{j\nu} = 0$. In fact (7) is invariant under the transformation $\theta_{i\nu} \to \theta_{i\nu} + c$ for all i, $\epsilon_{j\nu} \to \epsilon_{j\nu} - c$ for all j, where c is any real number.

In model (8) we consider now the problem of estimating the $\epsilon_{j\nu}$'s. Estimates of the ϵ's and ϕ's of (7) can then easily be derived from the estimates of $\epsilon_{j\nu}$'s of (8). A description of this last step and of a test for whether the model has the reduced parameter structure (7) can be found in Andersen (1969 pp. 23–24). In short, one can derive conditional maximum-likelihood (m.l.) estimates for $\epsilon_1, \ldots, \epsilon_k$ and $\phi_1, \ldots, \phi_m$ by maximizing (9) with respect to $\epsilon_{j\nu}$ under the side conditions $\epsilon_{j\nu} = \epsilon_j\ \phi_\nu$. We can remove all indeterminacies in the parametrization by putting $\phi_1 = 1$, $\phi_m = 0$ and $\sum_{j=1}^{k} \epsilon_j = 0$.

3. The Estimation Procedure

We shall base the estimation of the ϵ's on a conditional procedure suggested by Rasch (1960, 1961) and justified by Andersen (1970). We shall roughly sketch the procedure here. For details the reader is referred to Andersen (1970).

From the exponential form of (8) it follows that the set of vectors $(\mathbf{X}_{i.})$ form a set of minimal sufficient estimates for the matrix $\{\theta_{i\nu}\}$ $(i = 1, \ldots, N,\ \nu = 1, \ldots, m)$ when the ϵ's are assumed known. Hence the conditional distribution of the observations given the value $(\mathbf{x}_{i.})$ of $(\mathbf{X}_{i.})$ will be independent of the θ's. If we take this conditional distribution as our reference, we can carry out a maximum-likelihood estimation of the ϵ's that does not involve the θ's. We denote the thus defined estimators "conditional maximum-likelihood estimators" (c.m.l. estimators).

The reason for conditioning is that the model belongs to the important class of models having both structural and incidental parameters. As shown by Neyman and Scott (1948) the direct m.l. estimator for a structural parameter in the presence of incidental parameters does not need to be even consistent. In the present model the θ's are incidental since their number increases as the sample increases, that is as $N \to \infty$. In Andersen (1970) it was shown that c.m.l. estimators in contrast to direct m.l. estimators are consistent and asymptotically normally distributed under suitable regularity conditions and provided that the θ's do not behave oddly as $N \to \infty$.

It is a matter of simple algebra to verify that the conditional distribution required is given by

$$f[\{\mathbf{x}_{ij}\} \,|\, (\mathbf{x}_{i.})] = \exp\left(\sum_{j=1}^{k} \sum_{\nu=1}^{m} \epsilon_{j\nu} x_{.j}^{(\nu)}\right) \prod_{i=1}^{N} [\gamma_{\mathbf{x}_{i.}}(\{\epsilon_{j\nu}\})]^{-1}, \tag{9}$$

where

$$\gamma_{\mathbf{x}_{i.}}(\{\epsilon_{j\nu}\}) = \sum_{z} \exp\left(\sum_{j=1}^{k} \sum_{\nu=1}^{m} \epsilon_{j\nu} x_{ij}^{(\nu)}\right), \tag{10}$$

in which z stands for $\sum_{j=1}^{k} x_{ij}^{(\nu)} = x_{i.}^{(\nu)}$ $(\nu = 1, \ldots, m)$; that is, the expression

$$\exp\left(\sum_{j=1}^{k} \sum_{\nu=1}^{m} \epsilon_{j\nu} x_{ij}^{(\nu)}\right)$$

summed over all set of values of $x_{ij}^{(\nu)}$ $(j = 1, \ldots, k,\ \nu = 1, \ldots, m)$ that satisfy the condition $\sum_{j=1}^{k} x_{ij}^{(\nu)} = x_{i.}^{(\nu)}$ for $\nu = 1, \ldots, m$. In fact (9) is a rather direct version of the general formula for the conditional distributions given a minimal sufficient estimate found in textbooks as, for example, Lehmann (1959), Lemma 6, p. 47.

The c.m.l. estimate is now obtained by maximizing (9) with respect to the ϵ's under the side-conditions

$$\epsilon_{jm} = 0, \quad j = 1, \ldots, k, \tag{11}$$

and

$$\sum_{\nu=1}^{k} \epsilon_{j\nu} = 0, \quad \nu = 1, \ldots, m. \tag{12}$$

By putting the logarithmic derivatives of (9) with respect to the ϵ's equal to zero we get under (11) and (12) the (conditional) likelihood equations

$$x_{.p}^{(\mu)} - x_{.k}^{(\mu)} = \sum_{i=1}^{N} \partial \log \gamma_{\mathbf{x}_{i.}}(\{\epsilon_{j\nu}\}) \Big/ \partial \epsilon_{p\mu} - \sum_{i=1}^{N} \partial \log \gamma_{\mathbf{x}_{i.}}(\{\epsilon_{j\nu}\}) / \partial \epsilon_{k\mu},$$

$$p = 1, \ldots, k-1, \quad \mu = 1, \ldots, m-1. \tag{13}$$

That equation (13) has a unique set of solutions, which are the maximizing values of (9), follows from well-known results on exponential families of distributions, see, for example, Kendall and Stuart (1961, Vol. II pp. 52–53).

In order to solve (13) we shall in the next section study the γ-functions in more detail.

The likelihood equations for estimating $\epsilon_1, \ldots, \epsilon_k$ and $\phi_1, \ldots, \phi_m$ of the original model can easily be obtained from (13). In fact the new set of equations are simply weighted sums of (13) with $\epsilon_1, \ldots, \epsilon_k$ and $\phi_1, \ldots, \phi_m$ as weights as is easily seen by partial differentiations of (9) with respect to $\epsilon_{j\nu}$ under the side-conditions $\epsilon_{j\nu} = \epsilon_j \phi_\nu$ $(j = 1, \ldots, k; \nu = 1, \ldots, m)$.

4. The γ-functions

We start by simplifying the expression for $\gamma_{\mathbf{x}_i}(\{\epsilon_{j\nu}\})$. Instead of the ϵ's we introduce the transformed parameters

$$\delta_{j\nu} = \exp(\epsilon_{j\nu}) \tag{14}$$

in the expression for $\gamma_{\mathbf{x}_{i.}}(\{\epsilon_{j\nu}\})$. We note further that the possible values of the vector $\mathbf{x}_{i.} = (x_{i.}^{(1)}, \ldots, x_{i.}^{(m)})$ are $\mathbf{r} = (r_1, \ldots, r_m)$, where the r's are integers satisfying $\sum_{\nu=1}^{m} r_\nu = k$. With these notations the general form of a γ-function is

$$\gamma_{\mathbf{r}}(\{\delta_{j\nu}\}) = \sum_{i(1), \ldots, i(k)} \delta_{i(1).1} \cdots \delta_{i(r_1).1} \cdots \delta_{i(r_1+\ldots+r_{m-1}+1).m} \cdots \delta_{i(k).m}, \tag{15}$$

where $i(1), \ldots, i(k)$ is any permutation of the set $(1, \ldots, k)$ satisfying $i(1) < \ldots < i(r_1)$, $i(r_1+1) < \ldots < i(r_1+r_2), \ldots,\ \ i(r_1+\ldots+r_{m-1}+1) < \ldots < i(k)$. Thus each term in the summation consists of k factors, r_1 of the factors from $(\delta_{11}, \ldots, \delta_{k1})$, r_2 factors from $(\delta_{12}, \ldots, \delta_{k2})$, etc., with the restriction that $\delta_{j\nu}$ and $\delta_{j\mu}$ for the same j and $\nu \neq \mu$ cannot appear in the same term. The expression is slightly simplified by the fact that $\delta_{jm} = 1$ for all j according to (11) and (14). This means that the typical addend in (15) is the product of r_1 factors from $(\delta_{11}, \ldots, \delta_{k1}), \ldots, r_{m-1}$ factors from $(\delta_{1.m-1}, \ldots, \delta_{k.m-1})$. A consequence of this is that in $\gamma_r(\{\delta_{j\nu}\})$ we only need to specify $r_1, \ldots, r_{m-1}$. A simple combinational argument now shows that under the convention that $\gamma_{r-e_\mu}(\ldots) = 0$ if $r_\mu = 0$ in r we have the recursive formula

$$\gamma_{\mathbf{r}}(\{\delta_{j\nu}\}) = \gamma_{\mathbf{r}}(\{\delta_{j\nu}\}^{(p)}) + \sum_{\mu=1}^{m-1} \delta_{p\mu} \gamma_{\mathbf{r}-\mathbf{e}_\mu}(\{\delta_{j\nu}\}^{(p)}), \tag{16}$$

where $\{\delta_{j\nu}\}^{(p)}$ is defined as the matrix $\{\delta_{j\nu}\}$ with the pth row removed, and $\mathbf{r}-e_\mu = (r_1, \ldots, r_\mu - 1, \ldots, r_m)$. Formula (16) could also be derived from the fact that $\gamma_{\mathbf{r}}(\{\delta_{j\nu}\})$ is the coefficient of

$$(z_1 \ldots z_k)(u_1^{r_1} \ldots u_m^{r_m})$$

in

$$\prod_{j=1}^{k} \prod_{\mu=1}^{m} (1 + \delta_{j\mu} z_j u_\mu),$$

which can be re-written as

$$\left\{\prod_{\mu=1}^{m} (1 + \delta_{p\mu} z_p u_\mu)\right\} \left\{\prod_{j \neq p}^{k} \prod_{\nu=1}^{m} (1 + \delta_{j\nu} z_j u_\nu)\right\}.$$

It should be noted that in the first term of the right-hand side of (16), $\mathbf{r} = \mathbf{r} - e_m$, but since we do not have to specify the value of r_m, it is convenient to use the same symbol for $\mathbf{r}$, $\mathbf{r}-e_m$, $\mathbf{r}-e_m-e_m$, etc. Actually the value of r_m is determined by the number of rows in the δ-matrix from which the γ-function is derived.

Formula (16) defines all $\gamma_{\mathbf{r}}$'s recursively since we have the initial value

$$\gamma_{0,0,\ldots,0}(\{\delta_{j\nu}\}) = 1 \tag{17}$$

for all values of $\{\delta_{j\nu}\}$.

In Section 5 we shall show that the c.m.l. estimates are the roots of an equation, which contains the γ-functions through the ratios

$$\gamma_{\mathbf{r}-e_\mu}(\{\delta_{j\nu}\}^{(p)})/\gamma_{\mathbf{r}}(\{\delta_{j\nu}\}). \tag{18}$$

We shall start by showing, therefore, how (16) can be used to obtain a recursive procedure for the computation of the ratio (18). For the sake of brevity we write

$$\gamma_{\mathbf{r}-e_\mu}(\{\delta_{j\nu}\}^{(p)}) = \gamma_{\mathbf{r}-e_\mu}^{(p)}$$

and

$$\gamma_{\mathbf{r}}(\{\delta_{j\nu}\}) = \gamma_{\mathbf{r}}.$$

From (16) we get by dividing by $\gamma_{\mathbf{r}}$

$$\gamma_{\mathbf{r}}^{(p)}/\gamma_{\mathbf{r}} = 1 - \sum_{\mu=1}^{m-1} \delta_{p\mu} (\gamma_{\mathbf{r}-e_\mu}^{(p)}/\gamma_{\mathbf{r}-e_\mu})(\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}), \tag{19}$$

which gives us a recursive formula for $\gamma_{\mathbf{r}}^{(p)}/\gamma_{\mathbf{r}}$ together with the initial values

$$\gamma_{0,0,\ldots,0}^{(p)}/\gamma_{0,0,\ldots,0} = 1 \quad \text{for all } p, \tag{20}$$

if the quantities $\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}$ are available. We observe, however, that (16) applies to the functions $\gamma_{\mathbf{r}}^{(p)}$ as well. Thus

$$\gamma_{\mathbf{r}}^{(p)} = \gamma_{\mathbf{r}}^{(p,q)} + \sum_{\mu=1}^{m-1} \delta_{q\mu} \gamma_{\mathbf{r}-e_\mu}^{(p,q)}, \tag{21}$$

where $\gamma_{\mathbf{r}}^{(p,q)} = \gamma_{\mathbf{r}}(\{\delta_{j\nu}\}^{(p,q)})$ and $\{\delta_{j\nu}\}^{(p,q)}$ is the δ-matrix with both the pth and the qth row removed. Taking $p = 1$, $q = 2$ and repeating the argument we get the system of formulae

$$\gamma_{\mathbf{r}} = \gamma_{\mathbf{r}}^{(1)} + \sum_{\mu=1}^{m-1} \delta_{1\mu} \gamma_{\mathbf{r}-e_\mu}^{(1)}, \tag{22}$$

$$\gamma_{\mathbf{r}}^{(1)} = \gamma_{\mathbf{r}}^{(1,2)} + \sum_{\mu=1}^{m-1} \delta_{2\mu} \gamma_{\mathbf{r}-e_\mu}^{(1,2)}, \tag{23}$$

$$\gamma_{\mathbf{r}}^{(1,\ldots,j)} = \gamma_{\mathbf{r}}^{(1,\ldots,j+1)} + \sum_{\mu=1}^{m-1} \delta_{j+1,\mu} \gamma_{\mathbf{r}-e_\mu}^{(1,\ldots,j+1)}. \tag{24}$$

Actually the chain stops before $j = k$ unless $\mathbf{r} = (0, \ldots, 0)$. This is due to the fact that if $\sum_{\nu=1}^{m-1} r_\nu = c > 0$ then the addends of the right-hand side of (15) exist only if no more than $k-c$ rows are removed from $\{\delta_{j\nu}\}$. Therefore in (24), $j \leqslant k - \sum_{\nu=1}^{m-1} r_\nu$ and if $j = k - \sum_{\nu=1}^{m-1} r_\nu$ then $\gamma_{\mathbf{r}}^{(1,\ldots,j+1)} = 0$.

Suppose now that we divide the complete list of possible values of $\mathbf{r}$ in sections according to the value of $\sum_{\nu=1}^{m-1} r_\nu = c$ as shown in the first column of Table 1. Then $\mathbf{r} - e_\mu$ belongs to section $c-1$ and formulae (22)–(24) give $\gamma_{\mathbf{r}}$ as a function of the

TABLE 1

Computational scheme for $\gamma_{\mathbf{r}}$

$\mathbf{r}$	*Section code* c	$\gamma_{\mathbf{r}}$	$\gamma_{\mathbf{r}}^{(1)}$	...	$\gamma_{\mathbf{r}}^{(1,\ldots,k-c)}$	...	$\gamma_{\mathbf{r}}^{(1,\ldots,k)}$
0, 0, ..., 0	0	1	1	...	1	...	1
0, 0, ..., 1	1						—
...	...						—
0, 1, ..., 0	1						—
1, 0, ..., 0	1						—
...	...						
0, 0, ..., c	c					—	—
...	...					—	—
c, 0, ..., 0	c					—	—
...	...						
0, 0, ..., k	k		—	—	—	—	—
...	...		—	—	—	—	—
k, 0, ..., 0	k		—	—	—	—	—

δ's and $\gamma_{\mathbf{r}}$'s from section $c-1$. For $c = 0$, $\mathbf{r} = (0, \ldots, 0)$ is the only member of the section and $\gamma_{\mathbf{r}} = 1$, $\gamma_{\mathbf{r}}^{(1,\ldots,j)} = 1$, $j = 1, \ldots, k-1$. To fit the following formulae we put $\gamma_{\mathbf{r}}^{(1,\ldots,k)} = 1$. For $0 < c \leqslant k$ the starting point is $j = k-c$, where $\gamma_{\mathbf{r}}^{(1,\ldots,j)}$ reduces to

$$\gamma_{\mathbf{r}}^{(1,\ldots,k-c)} = \sum_{\mu=1}^{m-1} \delta_{k+1-c,\mu} \gamma_{\mathbf{r}-e_\mu}^{(1,\ldots,k+1-c)}. \tag{25}$$

$\gamma_{\mathbf{r}}^{(1,\ldots,j)}$ is then computed successively from (24) for $j = k-c-1, \ldots, j = 1$ and finally $\gamma_{\mathbf{r}}$ is obtained from (22). The computational scheme for this procedure is shown in

Table 1. Note that only $r_1, \ldots, r_{m-1}$ is listed in $\mathbf{r}$ in accordance with the notations introduced in (16).

The triangular character of Table 1 explained above is indicated in the table.

It is obvious from (15) that the $\gamma_{\mathbf{r}}$'s increase their values rather drastically with m and k. It is necessary therefore to introduce the auxiliary quantities $\eta_{\mathbf{r}}$ and $\eta_{\mathbf{r}}^{(1,\ldots,j)}$ that satisfy

$$\eta_{\mathbf{r}}^{(1,\ldots,j)} = \eta_{\mathbf{r}}^{(1,\ldots,j+1)}/m_{c-1} + \sum_{\mu=1}^{m-1} \delta_{j+1,\mu}\, \eta_{\mathbf{r}-e_\mu}^{(1,\ldots,j+1)}\, m_{c-1}, \tag{26}$$

where $m_c = \max\{\eta_{\mathbf{r}} \mid \sum_{\mu=1}^{m-1} r_\mu = c\}$. By comparing (24) and (26) it is readily seen that

$$\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}} = \eta_{\mathbf{r}-e_\mu}(m_{c-1}\,\eta_{\mathbf{r}})^{-1}. \tag{27}$$

Hence if we replace the $\gamma_{\mathbf{r}}$'s of Table 1 by the $\eta_{\mathbf{r}}$'s by introducing the norming constants $m_c\,(c = 0, \ldots, k)$, we can compute the quantities $\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}$ directly from (27) and the ratios $\gamma_{\mathbf{r}}^{(p)}/\gamma_{\mathbf{r}}$ recursively from (19) taking the $\mathbf{r}$'s in the order specified by column 1 of Table 1. Finally (18) is obtained as

$$\gamma_{\mathbf{r}-e_\mu}^{(p)}/\gamma_{\mathbf{r}} = (\gamma_{\mathbf{r}-e_\mu}^{(p)}/\gamma_{\mathbf{r}-e_\mu})(\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}). \tag{28}$$

That (19) is actually a recursive formula for $\gamma_{\mathbf{r}}^{(p)}/\gamma_{\mathbf{r}}$ follows from the fact that when $\mathbf{r}$ belongs to section c of the $\mathbf{r}$-list shown in column 1 of Table 1 then $\mathbf{r}-e_\mu$ belongs to section $c-1$ and hence $\mathbf{r}-e_\mu$ precedes $\mathbf{r}$ in the list.

5. The Solution of the Estimation Equation

Without the constraints $\sum_{j=1}^{k} \epsilon_{j\nu} = 0$ on the ϵ's we get immediately from (16) and (14) that

$$\partial\gamma_{\mathbf{r}}(\{\delta_{j\nu}\})/\partial\epsilon_{p\mu} = \delta_{p\mu}\,\gamma_{\mathbf{r}-e_\mu}(\{\delta_{j\nu}\}^{(p)}) = \delta_{p\mu}\,\gamma_{\mathbf{r}-e_\mu}^{(p)}$$

and hence

$$\partial \log \gamma_{\mathbf{x}_{i.}}(\{\epsilon_{j\nu}\})/\partial\epsilon_{p\mu} = \delta_{p\mu}\,\gamma_{\mathbf{x}_{i.}-e_\mu}^{(p)}/\gamma_{\mathbf{x}_{i.}}. \tag{29}$$

But, denoting by $n_{\mathbf{r}}$ the number of persons which have $\mathbf{x}_{i.} = \mathbf{r}$, we get from (13) and (29) the likelihood equations to be solved as

$$x_{.p}^{(\mu)} - x_{.k}^{(\mu)} = \delta_{p\mu} \sum_{\mathbf{r}} n_{\mathbf{r}}\, \gamma_{\mathbf{r}-e_\mu}^{(p)}/\gamma_{\mathbf{r}} - \delta_{k\mu} \sum_{\mathbf{r}} n_{\mathbf{r}}\, \gamma_{\mathbf{r}-e_\mu}^{(k)}/\gamma_{\mathbf{r}}, \quad \mu = 1, \ldots, m-1, \quad p = 1, \ldots, k-1. \tag{30}$$

We put

$$F_{p\mu} = \sum_{\mathbf{r}} n_{\mathbf{r}}\, \gamma_{\mathbf{r}-e_\mu}^{(p)}/\gamma_{\mathbf{r}}, \tag{31}$$

so that (30) becomes

$$x_{.p}^{(\mu)} - x_{.k}^{(\mu)} = \delta_{p\mu} F_{p\mu} - \delta_{k\mu} F_{k\mu}. \tag{32}$$

We want to apply Fisher's scoring method for the solution of (32). To this end we need the derivatives of (32) with respect to $\epsilon_{j\nu}$. From (16), (21) and (31) we get

$$\begin{aligned}\partial\{\delta_{p\mu} F_{p\mu} - \delta_{k\mu} F_{k\mu}\}/\partial\epsilon_{j\nu} &= \delta_{p\mu}\,\delta_{j\nu}\, C_{p\mu.j\nu} - \delta_{p\mu}\,\delta_{k\nu}\, C_{p\mu.k\nu} \\ &\quad - \delta_{k\mu}\,\delta_{j\nu}\, C_{k\mu.j\nu} + \delta_{k\mu}\,\delta_{k\nu}\, C_{k\mu.k\nu} = M_{p\mu.j\nu},\end{aligned}$$

where

$$C_{p\mu.p\mu} = F_{p\mu}/\delta_{p\mu} - \sum_{\mathbf{r}} n_{\mathbf{r}} (\gamma^{(p)}_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}})^2, \tag{33}$$

$$C_{p\mu.p\nu} = -\sum_{\mathbf{r}} n_{\mathbf{r}} (\gamma^{(p)}_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}) (\gamma^{(p)}_{\mathbf{r}-e_\nu}/\gamma_{\mathbf{r}}), \quad \text{for } \nu \neq \mu, \tag{33'}$$

and

$$C_{p\mu.j\nu} = \sum_{\mathbf{r}} n_{\mathbf{r}} (\gamma^{(pj)}_{\mathbf{r}-e_\mu-e_\nu}/\gamma_{\mathbf{r}}) - \sum_{\mathbf{r}} n_{\mathbf{r}} (\gamma^{(p)}_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}) (\gamma^{(j)}_{\mathbf{r}-e_\nu}/\gamma_{\mathbf{r}}), \quad \text{for } p \neq j \text{ and all } \nu, \mu. \tag{33''}$$

In these expressions there appear, in addition to terms of the form $\gamma^{(p)}_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}$ terms of the form $\gamma^{(pj)}_{\mathbf{r}-e_\mu-e_\nu}/\gamma_{\mathbf{r}}$. These can be computed recursively, however, from (21), which can be rewritten as

$$\gamma^{(pj)}_{\mathbf{r}}/\gamma_{\mathbf{r}} = \gamma^{(p)}_{\mathbf{r}}/\gamma_{\mathbf{r}} - \sum_{\mu=1}^{m-1} \delta_{j\mu} (\gamma^{(pj)}_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}-e_\mu}) (\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}), \tag{34}$$

where $\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}$ has been computed in Table 1 and $\gamma^{(p)}_{\mathbf{r}}/\gamma_{\mathbf{r}}$ is given by (19). The recursive formula (34) works in the same way as (19) starting with $\gamma^{(pj)}_{0,0,\ldots,0} = 1$ and proceeding through the complete list of $\mathbf{r}$'s shown in column 1 of Table 1. We now have

$$\gamma^{(pj)}_{\mathbf{r}-e_\mu-e_\nu}/\gamma_{\mathbf{r}} = (\gamma^{(pj)}_{\mathbf{r}-e_\mu-e_\nu}/\gamma_{\mathbf{r}-e_\mu-e_\nu}) (\gamma_{\mathbf{r}-e_\mu-e_\nu}/\gamma_{\mathbf{r}-e_\mu}) (\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}). \tag{35}$$

While $\gamma^{(p)}_{\mathbf{r}}/\gamma_{\mathbf{r}} = 0$ in the last section of the $\mathbf{r}$-list, $\gamma^{(pj)}_{\mathbf{r}}/\gamma_{\mathbf{r}} = 0$ in both section $k-1$ and section k.

The estimation procedure is now as follows. An initial matrix of ϵ's is chosen $\{\epsilon^0_{j\nu}\}$, say. From Table 1 the ratios $\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}$ are computed with $\delta_{j\nu} = \exp(\epsilon^0_{j\nu})$. From (19) and (28) we determine $\gamma^{(p)}_{\mathbf{r}-e_\mu}\gamma_{\mathbf{r}}$ for all $\mathbf{r}$, μ and p and from (34) and (35) we determine $\gamma^{(pj)}_{\mathbf{r}-e_\mu-e_\nu}$ for all $\mathbf{r}$, μ, ν, p and j. It is then straightforward to compute the quantities $F_{p\mu}$ and $C_{p\mu.j\nu}$ for all p, j, μ and ν from (31) and (33).

Since the C's are derivatives of the likelihood equation (32) to be solved we have according to Fisher's scoring method the following system of equations to determine an improved set of ϵ-estimates:

$$\sum_{\nu=1}^{m-1} \sum_{j=1}^{k-1} (\epsilon_{j\nu} - \epsilon^0_{j\nu}) (\delta^0_{p\mu}\, \delta^0_{j\nu}\, C_{p\mu.j\nu} - \delta^0_{p\mu}\, \delta^0_{k\nu}\, C_{p\mu.k\nu} - \delta^0_{k\mu}\, \delta^0_{j\nu}\, C_{k\mu.j\nu} + \delta^0_{k\mu}\, \delta^0_{k\nu}\, C_{k\mu.k\nu})$$

$$= x^{(\mu)}_{.p} - x^{(\mu)}_{.k} - \delta^0_{p\mu} F_{p\mu} + \delta^0_{k\mu} F_{k\mu}, \tag{36}$$

where we recall that $\delta^0_{p\mu} = \exp(\epsilon^0_{p\mu})$. We can repeat the procedure taking the improved $\epsilon_{j\nu}$ as $\epsilon^0_{j\nu}$ and we stop the iterative procedure when the two sides of (32) are satisfactorily close to zero.

As a by-product of the estimation procedure we get estimates of the uncertainty of the estimates, since the inverse of the matrix $\{M_{p\mu.j\nu}\}$ estimates the asymptotic variance matrix of the estimates. This follows from the main theorem of Andersen (1970).

In the next section we shall illustrate the developed estimation method by a numerical example.

6. A NUMERICAL EXAMPLE

We consider a questionnaire containing 4 questions each with 3 possible answers. The data consists of 300 persons answers to the 4 questions. In Table 2 the quantities $x_{.p}^{(\mu)}$ and $n_{\mathbf{r}}$ appearing in the estimation equation (30) are shown for this set of data.

TABLE 2

Question totals and person group numbers for 300 persons answers to 4 questions

Question totals

Question	*Answer category* 1	2	3
1	169	94	37
2	135	110	55
3	63	88	149
4	59	98	143

Person group numbers

$\mathbf{r}$:	00	01	10	02	11	20	03	12	21	30	04	13	22	31	40
$n_{\mathbf{r}}$:	6	28	14	25	33	15	12	32	32	12	4	19	20	34	14

These data were simulated by a computer with the θ's being a random sample of 300 observations from a normal distribution with mean 0 and variance 1 and the ϵ's being equal to

$\epsilon_{j\nu}$	$\nu = 1$	2	3
$j = 1$	2·0	1·0	0·0
2	1·0	0·5	0·0
3	−1·4	−0·7	0·0
4	−1·6	−0·8	0·0

Examples of the estimation method applied to real data are reported by Andersen (1969). The simulated data were chosen to have k and m small enough to be able to display the complete matrices given by Tables 3–6. The data of Andersen (1969) are with $k = 8$ and $m = 4$ and do not permit this.

Initial values for the estimation are constructed from (30). If we put all $\delta_{i\mu}$ $(i = 1, \ldots, k)$ equal to the same values (the common value is 1 by (11) and (14)), (30) reduces to

$$x_{.p}^{(\mu)} - x_{.k}^{(\mu)} = (\delta_{p\mu} - \delta_{k\mu}) A_{\mu}, \tag{37}$$

where A_{μ} is a constant which is independent of p, or

$$x_{.p}^{(\mu)} - \delta_{p\mu} A_{\mu} = C_{\mu}, \quad \text{for } p = 1, \ldots, k, \tag{38}$$

where C_μ is independent of p. Since it can be shown that

$$\sum_{p=1}^{k} x_{.p}^{(\mu)} = \sum_{p=1}^{k} \delta_{p\mu} \sum_{\mathbf{r}} n_{\mathbf{r}} \gamma_{\mathbf{r}-e_\mu}^{(p)}/\gamma_{\mathbf{r}},$$

it follows from (30) and (37) that we may put $C_\mu = 0$ for all μ. Hence (38) reduces to

$$x_{.p}^{(\mu)} = A_\mu \, \delta_{p\mu}, \quad p = 1, \ldots, k. \tag{39}$$

The constraints (11) and (12) imply that $\delta_{j\mu} = \exp(\epsilon_{j\mu})$ must satisfy $\delta_{jm} = 1$ for all j and $\prod_{j=1}^{k} \delta_{j\mu} = 1$ for all μ. From (39) it follows consequently that we can take

$$\delta_{p\mu}^{(0)} = (x_{.p}^{(\mu)}/x_{.p}^{(m)}) \left\{ \prod_{j=1}^{k} (x_{.j}^{(m)}/x_{.j}^{(\mu)}) \right\}^{1/k}, \quad p = 1, \ldots, k, \quad \mu = 1, \ldots, m, \tag{40}$$

as initial δ-values.

For the data from Table 2 the initial δ-values are

$$\{\delta_{j\nu}^{0}\} = \begin{bmatrix} 3.86 & 2.12 & 1.00 \\ 2.08 & 1.67 & 1.00 \\ 0.36 & 0.49 & 1.00 \\ 0.35 & 0.57 & 1.00 \end{bmatrix}. \tag{41}$$

TABLE 3

Computed values of $\eta_{\mathbf{r}}^{(1,\ldots,j)}$ *based on the* δ*'s given by* (41)

r	$j = 0$	1	2	3	4	m_c
00	1·00	1·00	1·00	1·00	1·00	
01	4·857	2·736	1·066	0·572		6·64
10	6·644	2·782	0·706	0·349		
02	1·184	0·310	0·042			3·05
11	3·046	0·567	0·057			
20	1·857	0·239	0·019			
03	0·239	0·023				0·94
12	0·849	0·060				
21	0·935	0·049				
30	0·316	0·013				
04	0·053					
13	0·232					
22	0·359					
31	0·231					
40	0·053					

The computation of $\eta_{\mathbf{r}}^{(1,\ldots,j)}$ from (26) based on these initial parameters is shown in Table 3.

From the numbers in Table 3, we can subsequently compute $\gamma_{\mathbf{r}-e_\mu}/\gamma_{\mathbf{r}}$ using (27), $\gamma_{\mathbf{r}}^{(p)}/\gamma_{\mathbf{r}}$ using (19) and $\gamma_{\mathbf{r}}^{(pj)}/\gamma_{\mathbf{r}}$ using (34). In order to save space the actual figures are not shown. In Table 4 we have shown the computation of $F_{p\mu}$ given by (31) for $p = 1$ and $\mu = 1$ to give an impression of the order of magnitude of the γ-ratios.

TABLE 4

Computed values of $\gamma_{\mathbf{r}-e_1}^{(1)}/\gamma_{\mathbf{r}}$ and F_{11}

	$\gamma_{\mathbf{r}-e_1}^{(1)}/\gamma_{\mathbf{r}}$
10	0·150
11	0·135
20	0·226
12	0·120
21	0·199
30	0·248
13	0·107
22	0·179
31	0·225
40	0·256
F_{11}	40·08

Table 5 shows the ϵ-estimates after one iteration and after the final iteration with the matrix (41) as initial values. The required accuracy is 5×10^{-4}.

TABLE 5

The estimated parameters after first and final iteration for an example with $m = 3$, $k = 4$

Iteration	ϵ_{ij}:	$j = 1$	2	3
1	$i = 1$	1·351	0·752	0·000
	2	0·730	0·513	0·000
	3	−1·028	−0·707	0·000
	4	−1·053	−0·558	0·000
4	1	1·887	0·999	0·000
	2	1·064	0·680	0·000
	3	−1·490	−0·947	0·000
	4	−1·461	−0·732	0·000

It may also be of some interest to see what the estimate mentioned in Section 5 of the asymptotic covariance looks like. Table 6 shows M^{-1}.

We note that the standard error of the estimates is of order 0.17. Hence we can expect the true values to lie within the estimates ± 0.35. A comparison with the true parameter values shows that this is actually the case.

TABLE 6

Asymptotic covariance matrix of c.m.l. estimates

$M^{-1}_{j\mu.p\nu}$	$p = 1$		$p = 2$		$p = 3$	
	$\nu = 1$	2	$\nu = 1$	2	$\nu = 1$	2
$j = 1$ $\mu = 1$	0·035	0·023	−0·006	−0·006	−0·014	−0·009
2	0·023	0·029	−0·006	−0·007	−0·009	−0·011
$j = 2$ $\mu = 1$	−0·006	−0·006	0·029	0·017	−0·011	−0·006
2	−0·006	−0·007	0·017	0·023	−0·006	−0·008
$j = 3$ $\mu = 1$	−0·014	−0·009	−0·011	−0·006	0·030	0·013
2	−0·009	−0·011	−0·006	−0·008	0·013	0·020

ACKNOWLEDGEMENTS

A considerable part of the program writing and almost all the program testing and handling of the numerical examples was done by Mr Peter Henningsen. The computations were carried out on a IBM 7094 at the Northern Europe University Computing Center.

REFERENCES

ANDERSEN, E. B. (1964). *The Discrete Measurement Model.* (In Danish.) Copenhagen: University of Copenhagen Press.

—— (1969). Conditional inference and multiple choice questionnaires. Technical Report. Copenhagen School of Economics.

—— (1970). Asymptotic properties of conditional maximum likelihood estimates. *J. R. Statist. Soc.* B, **32**, 283–301.

KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics*, Vol. II. London: Griffin.

LEHMANN, E. (1959). *Testing Statistical Hypotheses.* New York: Wiley.

PETERSEN, E. (1968). *Thriving in Danish Factories.* (In Danish.) Copenhagen: Mentalhygiejnisk Institut.

RASCH, G. (1960). *Some Models for Intelligence and Attainment Tests.* Copenhagen: Teknisk Forlag.

—— (1961). On general laws and the meaning of measurement in psychology. In *Proc. Fourth Berk. Symp.*, Vol. IV, pp. 321–333.