Open access • Journal Article • DOI:10.1177/014662168200600406

Latent trait models and ability parameter estimation — Source link

Erling B. Andersen

Institutions: University of Copenhagen

Published on: 01 Sep 1982 - Applied Psychological Measurement (SAGE Publications)

Topics: Latent variable model, Latent class model, Goodness of fit, Contingency table and Estimation theory

Latent Trait Models and Ability Parameter Estimation

Erling B. Andersen
University of Copenhagen

In recent years several authors have viewed latent trait models for binary data as special models for contingency tables. This connection to contingency table analysis is used as the basis for a survey of various latent trait models. This article discusses estimation of item parameters by conditional, direct, and marginal maximum likelihood methods, and estimation of individual latent parameters as opposed to an estimation of the parameters of a latent population density. Various methods for testing the goodness of fit of the model are also described. Several of the estimators and tests are applied to a data set concerning consumer complaint behavior.

During the last 20 years there has been a considerable amount of research on latent trait models of the logistic form. Since 1965 the connection to modern theory for exponential family distributions has been recognized and has been widely used; and in recent years the connection to various forms of contingency table analysis and latent structure analysis has been noted.

It is not the purpose of the present article to survey all of these developments, as there are a number of excellent survey papers available (Baker, 1977; Hambleton & Cook, 1977; Hambleton, Swaminathan, Cook, Eigner, & Gifford, 1978; Wainer, Morgan, & Gustavsson, 1980; Weiss & Davison, 1981) and text books (Andersen, 1980b; Bock, 1975; Lord, 1981; Lord & Novick, 1968; Wright & Stone, 1979) that deal with latent trait models. Rather, the aim is to give a short account of available inference methods for a logistic latent trait model and to assess the potentials of the methods. In pursuing the latter it will be shown how the various approaches apply to a new set of data.

Latent Trait Models and Contingency Tables

It has been increasingly the practice to write a latent trait model as a contingency table, allowing users to see the connection to a contingency table analysis approach to the data. In the present article this practice will be followed by writing the data from a test of k dichotomous items as a $2 \times 2 \ldots \times 2$ contingency table. For simplicity the situation will first be described for the $2 \times 2 \times 2 \times 2$ case. This corresponds to N individuals answering four items. The observed numbers in a $2 \times 2 \times 2 \times 2$ contingency table will then correspond to the number of observed response patterns on the four items.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 6, No. 4, Fall 1982, pp. 445–461

0416-6216/82/040445-17$1.85

In the following x_{ijkr} will be written for the cell count of cell (i, j, k, r) of the four-dimensional contingency table and p_{ijkr} for the corresponding cell probability. For two levels on each variable, a response pattern will then be a combination of four integers such as 1121 or 1222, if $i = 1, 2$, $j = 1, 2$, $k = 1, 2$, and $r = 1, 2$ are the indices of the observed levels. In item response theory it is customary, however, to use the indices 0 and 1 rather than 1 and 2, and this tradition will be followed. The possible response patterns are accordingly of the form 1101 and so forth or, in general, $ijkr$ with $i = 1, 0$, $j = 1, 0$, $k = 1, 0$, and $r = 1, 0$.

In contingency table analysis, the cell probabilities or the corresponding expected numbers in the cells are expressed in the form of a log-linear model, i.e.,

$$\ln p_{ijkr} = \tau_0 + \tau_i^{(1)} + \tau_j^{(2)} + \tau_k^{(3)} + \tau_r^{(4)} + \tau_{ij}^{(12)} + \tau_{ik}^{(13)} + \tau_{ir}^{(14)} + \tau_{jk}^{(23)} + \tau_{jr}^{(24)} + \tau_{kr}^{(34)} + \tau_{ijk}^{(123)} + \tau_{ijr}^{(124)} + \tau_{ikr}^{(134)} + \tau_{jkr}^{(234)} + \tau_{ijkr}^{(1234)} . \qquad [1]$$

All hypotheses that can be tested within the framework of log-linear models will take the form of collections of τ's being zero. In particular, all hypotheses of independence, conditional independence, and equal probabilities can be expressed as collections of τ's being zero.

Log-linear models are exponential family type models. A powerful statistical inference theory is available for analyzing such models (cf. Andersen, 1980b; Barndorff-Nielsen, 1978). For theory and applications of log-linear models for contingency tables, the reader should consult one of the many textbooks and monographs now available, as for example, Bishop, Fienberg, and Holland (1975), Haberman (1978), or Goodman (1978).

In many applications, however, the problem at hand is not readily expressed in terms of log-linear interaction parameters and hypotheses concerning such parameters. Very often what is looked for, instead, is some kind of joint link between the variables provided by some unobservable individual characteristic called a latent variable. Thus, for a known value of the latent variable the probability structure of the contingency table is simple, and the apparent dependencies between variables can be accounted for by the joint dependency on the latent variable.

For a 2 × 2 × 2 × 2 table the independence hypothesis is thus equivalent to the log-linear form

$$\ln p_{ijkr} = \tau_0 + \tau_i^{(1)} + \tau_j^{(2)} + \tau_k^{(3)} + \tau_r^{(4)} \qquad [2]$$

i.e., a model with no two-factor or higher order interactions. Suppose that this hypothesis is clearly rejected based on the observations at hand but that the problem underlying the data suggests that the dependency can be contributed to the variation of an unobservable latent individual parameter θ. It may, for example, be assumed that there is independence in the table, given the value of θ. This form of independence, which is often termed *local independence* in psychometric language, has the form

$$p_{ijkr}(\theta) = p_{i...}(\theta) p_{.j..}(\theta) p_{..k.}(\theta) p_{...r}(\theta) \qquad [3]$$

where $p_{ijkr}(\theta)$ is the probability of falling in cell (i,j,k,r) given the value of θ. Otherwise expressed, Equation 3 is the probability of falling in cell $(i,j,k,r,)$ for an individual with latent parameter θ.

The marginal probability is now obtained from Equation 3 as

$$p_{ijkr} = \int p_{ijkr}(\theta)\ \varphi(\theta)\,d\theta \ , \qquad [4]$$

where $\varphi(\theta)$ is the population density of θ.

The connection to classical psychometric modeling is now clear by observing that in the $2\times2\times2\times2$ case $p_{1\ldots}(\theta) = 1 - p_{0\ldots}(\theta)$, $p_{.1\ldots}(\theta) = 1 - p_{.0\ldots}(\theta)$ and so forth, so that the probabilities may be renamed as

$$p^{(1)}(\theta) = p_{1\ldots}(\theta) \ ,$$

$$p^{(2)}(\theta) = p_{.1..}(\theta) \ , \qquad [5]$$

and so forth, where $p^{(q)}(\theta) = P$ (answer 1 on item q given θ).

The model can now be expressed in the following form

$$p_{ijkr} = \int \left[p^{(1)}(\theta)\right]^{u_1}\left[1-p^{(1)}(\theta)\right]^{1-u_1} \cdots \left[p^{(4)}(\theta)\right]^{u_4} \left[1-p^{(4)}(\theta)\right]^{1-u_4} \varphi(\theta)\,d\theta \ , \qquad [6]$$

where $u_1, \ldots, u_4$ are binary response variables with values 0 or 1.

Without any further assumptions on $p^{(q)}(\theta)$ or $\varphi(\theta)$, an analysis based on Equation 6 would not be possible. In the literature there are two principal ways of dealing with Equation 6. One approach, known as latent class analysis, makes no assumptions as to the exact form of $p^{(1)}(\theta)$, $p^{(2)}(\theta)$, $p^{(3)}(\theta)$, and $p^{(4)}(\theta)$ but then assumes that $\varphi(\theta)$ is a discrete M-point distribution, i.e., for certain (unknown) cutting points on the θ-scale, $\theta_0, \ldots, \theta_M$,

$$\varphi_\nu = P(\theta_{\nu-1} < \theta < \theta_\nu) \ , \quad \nu = 1, \ldots, M \ . \qquad [7]$$

In latent class analysis this then results in the model

$$p_{ijkr} = \sum_{\nu=1}^{M} (p_\nu^{(1)})^{u_1}(1-p_\nu^{(1)})^{1-u_1} \cdots (p_\nu^{(4)})^{u_4} (1-p_\nu^{(4)})^{1-u_4}\ \varphi_\nu \ , \qquad [8]$$

where $p_\nu^{(q)} = p^{(q)}(\theta)$ is assumed constant for $\theta_{\nu-1} < \theta < \theta_\nu$. It has only $5M - 1$ parameters as compared with 15 p_{ijkr}'s. Hence, a model with $M = 2$ or $M = 3$ can be analyzed, when the contingency table is four-dimensional, while for higher dimensional tables, more class probabilities can be estimated. (Discussion of latent class analysis will be resumed in a later section.)

Another approach is to make few or no assumptions about $\varphi(\theta)$ but then to assume a certain structure in the $p^{(q)}(\theta)$'s. This approach is usually termed latent trait analysis, as the analysis is concerned with models for the probabilities $p^{(q)}(\theta)$. If, as in psychometrics, the variables are denoted as items, the probabilities are usually modeled in terms of certain item parameters.

As a very general form

$$p^{(q)}(\theta) = p(\theta, a_{q1}, \ldots, a_{qt}) \qquad [9]$$

where $a_{q1}, \ldots, a_{qt}$ are the parameters of item q.

A less general form, which covers most latent trait models for dichotomous items with a one-dimensional latent variable is

$$p^{(q)}(\theta) = g_0 + g_1 H\left[(\theta - b_q)a_q\right] , \qquad [10]$$

where $H(x)$ is an increasing function over the real line with $H(-\infty) = 0$ and $H(+\infty) = 1$. The so-called normal ogive model and the three-parameter logistic model can be written as in Equation 10 when g_0 is the guessing parameter, $g_1 = 1 - g_0$, $H(0) = 0.5$, b_q is the item difficulty, and a_q the item discriminating power. In the normal ogive model, H is the cumulative normal distribution function, and in the logistic model, H is the function $e^x/(1 + e^x)$. A very general discussion of latent trait models based on a formula similar to Equation 10 is given in a recent paper by Bartholomew (1980).

When there is a probability structure as in Equation 10, there is the possibility of assuming a certain form of the latent density $\varphi(\theta)$ including some population parameters. A model where $\varphi(\theta)$ is a normal density with mean μ and variance σ^2 has been discussed by Andersen and Madsen (1977) and Sanathanan and Blumenthal (1978) for a logistic model, and by Christofferson (1975) and Muthén (1978, 1979) for a normal ogive model. Bartholomew (1980) has discussed these and other models.

If $\varphi(\theta)$ depends on two parameters μ and σ, this gives the model

$$p_{ijkr} = \int \left[p^{(1)}(\theta)\right]^{u_1} \cdots \left[1 - p^{(4)}(\theta)\right]^{1-u_4} \varphi(\theta \mid \mu, \sigma^2)\, d\theta \qquad [11]$$

which, if the form of Equation 10 is assumed, will depend on the following parameters g_0, a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4, μ, and σ^2. With fifteen free p-parameters this is close to a reparameterization, but with more items, the parameters can be estimated and the model checked, although the likelihood function in most cases will have a complicated structure that gives rise to numerical problems in solving the likelihood equations.

The various models and statistical methods of this paper will be illustrated by a set of data derived from a Danish study of consumer complaint behavior. Six hundred individuals were faced with six typical consumer situations and were asked to state whether, under the given circumstances, they would complain or not. The basic data in terms of number of individuals N_u responding with each of the 64 possible response patterns u are shown in Table 1.

Estimation in the Logistic Latent Trait Model

This section will deal only with the logistic model[1], beginning with the one-parameter (Rasch) model, where in Equation 10 $g_0 = 0$, $g_1 = 1$, $H(x) = e^x/(1 + e^x)$ and $a_q = 1$, $q = 1, \ldots, k$. With these specifications the model is

$$p^{(q)}(\theta) = \exp(\theta - b_q) \;/\; [1 + \exp(\theta - b_q)] . \qquad [12]$$

In this section no assumption will be made on the latent density, returning to the estimation of $\varphi(\theta)$ and $\varphi_1, \ldots, \varphi_M$ in later sections of this article. Hence, the probability distribution for each individual, given the latent parameter θ, is considered. To this end is introduced

$$u_{sq} = \begin{cases} 1 \text{ if individual } s \text{ chooses category } 1 \text{ on item } q \\ 0 \text{ if individual } s \text{ chooses category } 0 . \end{cases} \qquad [13]$$

[1]Estimation in the normal ogive latent trait model is treated at length in Lord and Novick (1968). Readers are also referred to Bock (1972) or Muthen (1978).

Table 1
Number of Individuals for Each Response Pattern for Complaint Data

Response Pattern						N_u	Response Pattern						N_u
1	1	1	1	1	1	127	0	1	1	1	1	1	–
1	1	1	1	1	0	78	0	1	1	1	1	0	2
1	1	1	1	0	1	16	0	1	1	1	0	1	1
1	1	1	1	0	0	54	0	1	1	1	0	0	1
1	1	1	0	1	1	36	0	1	1	0	1	1	1
1	1	1	0	1	0	36	0	1	1	0	1	0	2
1	1	1	0	0	1	14	0	1	1	0	0	1	3
1	1	1	0	0	0	59	0	1	1	0	0	0	–
1	1	0	1	1	1	8	0	1	0	1	1	1	–
1	1	0	1	1	0	16	0	1	0	1	1	0	–
1	1	0	1	0	1	4	0	1	0	1	0	1	1
1	1	0	1	0	0	18	0	1	0	1	0	0	1
1	1	0	0	1	1	1	0	1	0	0	1	1	–
1	1	0	0	1	0	6	0	1	0	0	1	0	–
1	1	0	0	0	1	–	0	1	0	0	0	1	–
1	1	0	0	0	0	18	0	1	0	0	0	0	2
1	0	1	1	1	1	12	0	0	1	1	1	1	–
1	0	1	1	1	0	7	0	0	1	1	1	0	–
1	0	1	1	0	1	4	0	0	1	1	0	1	–
1	0	1	1	0	0	18	0	0	1	1	0	0	–
1	0	1	0	1	1	1	0	0	1	0	1	1	–
1	0	1	0	1	0	6	0	0	1	0	1	0	1
1	0	1	0	0	1	4	0	0	1	0	0	1	–
1	0	1	0	0	0	9	0	0	1	0	0	0	1
1	0	0	1	1	1	1	0	0	0	1	1	1	–
1	0	0	1	1	0	4	0	0	0	1	1	0	–
1	0	0	1	0	1	1	0	0	0	1	0	1	–
1	0	0	1	0	0	10	0	0	0	1	0	0	–
1	0	0	0	1	1	1	0	0	0	0	1	1	–
1	0	0	0	1	0	3	0	0	0	0	1	0	1
1	0	0	0	0	1	1	0	0	0	0	0	1	–
1	0	0	0	0	0	3	0	0	0	0	0	0	7

The probability of a response vector will then be

$$f(u_{s1}, \ldots, u_{sk}) = \prod_{q=1}^{k} \left[e^{(\theta_s - b_q)u_{sq}} \Big/ \left(1+e^{(\theta_s - b_q)}\right) \right]$$

$$= \exp\left(\theta_s u_{s.} - \sum_{q=1}^{k} b_q u_{sq}\right) \Big/ \prod_{q=1}^{k} \left(1+e^{(\theta_s - b_q)}\right) \qquad [14]$$

where θ_s is the latent parameter. From this equation it follows that the estimation of θ_s will depend only on the total score $u_{s.}$ and, from general results on exponential family distributions, that the maximum likelihood estimate of θ_s is the solution to

$$u_{s.} = E[U_{s.}] = \sum_{q=1}^{k} e^{(\theta_s - b_q)} \Big/ \left(1 + e^{(\theta_s - b_q)}\right) . \qquad [15]$$

For given values of b_q, this equation is easily solved by a Newton-Raphson method. (Details are given in Andersen, 1980b, chap. 6.)

In another application of exponential family theory, (cf. Andersen, 1980b, chaps. 3 and 6), the conditional probability of the response pattern $(u_{s1}, ..., u_{sk})$ given $u_{s.} = t$, say,

$$f(u_{s1}, \ldots, u_{sk} | u_{s.}) , \qquad [16]$$

is independent of the value of θ_s, and thus only depends on the b_q. This property of the Rasch model can be used to obtain conditional maximum likelihood (CML) estimates for $(b_1, ..., b_k)$[2].

When the CML approach is used for estimating the b_q's, the θ's are usually estimated from Equation 15 with the CML estimates for $b_1, ..., b_k$ inserted in the equation. The conditional likelihood equations are slightly involved, but various programs are available. Recently, the author's original program has been implemented as an SAS subroutine (cf. Weinreich, 1980).

There are other possible ways of estimating the b_q's. An alternative to the CML approach is a direct maximum likelihood (ML) approach. The direct likelihood will be the product of Equation 14 over all s, which takes the following form:

$$f(u_{11}, \ldots, u_{Nk}) = \exp\left(\sum_{s=1}^{N} \theta_s u_{s.} - \sum_{q=1}^{k} b_q u_{.q}\right) \Big/ \prod_s \prod_q \left(1 + e^{(\theta_s - b_q)}\right) \qquad [17]$$

Once again, from exponential family theory it can be deduced that the likelihood equations for the ML estimation become Equation 15 and, in addition,

$$u_{.q} = E[U_{.q}] = \sum_{s=1}^{N} e^{(\theta_s - b_q)} \Big/ \left(1 + e^{(\theta_s - b_q)}\right) . \qquad [18]$$

Note that Equation 15 has only $k - 1$ possible solutions, as all individuals with common value $u_{s.} = t$ get the same parameter estimate $\hat{\theta}(t)$ for $t = 1, ..., k - 1$. For the extreme values of t, $\hat{\theta}(0) = -\infty$ and $\hat{\theta}(k) = +\infty$ are obtained. The solutions to Equations 15 and 18 form a set of joint estimates $\hat{b}_1, ..., \hat{b}_k$, $\hat{\theta}(1), ..., \hat{\theta}(k - 1)$. These estimates for b_q do not coincide with the CML estimates. It can be shown that the ML estimate has a bias of $k/(k - 1)$, which does not vanish for large values of N. For large values of both k and N the bias is negligible and, as shown by Haberman (1977), the joint set of $\hat{b}_q$'s and $\hat{\theta}(t)$'s are consistent under certain restrictions on the joint convergence of k and N to infinity. The first computer program for the solution of Equations 15 and 18 is described by Lord (1968). The estimates can also be obtained by more general algorithms for exponential family distributions such as, for example, those in the GLIM library.

A comparison of the CML approach, the direct ML approach, and certain modifications of these

[2]In fact, it was this possibility of an individual parameter estimation independent of the b_q's that lead Rasch to suggest his model in 1960. Readers interested in the history of the Rasch model are referred to the reprinting of the 1960 book with comments by Wright (Rasch, 1980).

procedures is given by Wright and Douglas (1977). Approximate solutions for one of these modifications was given recently by Cohen (1979).

In a recent paper by Bock and Aitkin (1981) an alternative method called marginal maximum likelihood estimation is suggested. It is based on the EM algorithm and applies an approximation by Gauss-Hermite quadrature to the integral in Equation 6. By rescaling, $\varphi(\theta)$ is assumed to be a standard normal density, but the model can also by interpreted as using an empirical discrete latent distribution estimated directly. (For further details the reader is referred to the original paper.)

Comparison of Estimates

Some of these estimates for the data of Table 1 will now be examined. The item marginals and individual marginals are presented in Table 2. The CML estimates and their standard errors, and the corresponding solutions to Equation 18, are shown in Table 3. These estimates can be compared with the direct ML estimates (and their standard errors), which are given in Table 4. Note how clearly the multiplicative bias of $k/(k-1)$ emerges.

Table 2
Item and Individual Marginals for the Complaint Data

Item	Total	Score	Total
1	576	0	7
2	505	1	7
3	493	2	43
4	384	3	124
5	350	4	142
6	237	5	150
		6	127

Table 3
CML Estimates for Item Parameters and ML Estimates for Individual Parameters for Complaint Data

Item	$\hat{b}_q$	SE	Score	$\hat{\theta}(t)$	SE
1	−2.47	.21	0	$-\infty$	
2	−.64	.11	1	−2.20	1.27
3	−.48	.11	2	−.92	1.03
4	.66	.10	3	.06	.97
5	.96	.10	4	1.01	1.00
6	1.97	.10	5	2.16	1.20
			6	$+\infty$	

Table 4
Simultaneous ML Estimates for Both Item Parameters and Individual Parameters for the Complaint Data

Item	$\hat{b}_q$	SE	Score	$\hat{\theta}(t)$	SE
1	−2.95	.18	1	−2.45	1.33
2	−.84	.12	2	−1.03	1.09
3	−.64	.12	3	0.07	1.02
4	.77	.10	4	1.13	1.05
5	1.15	.10	5	2.40	1.25
6	2.52	.11			

Two-Parameter Model

Alternatively, it may be assumed that the two-parameter logistic model describes the data. The model is now given by

$$p^{(q)}(\theta) = \exp[(\theta-b_q)a_q] \; / \; \{1+\exp[(\theta-b_q)a_q]\} \;, \qquad [19]$$

where $a_1, ..., a_k$ are item discriminations. The probability of a given response vector $y_{s1}, ..., y_{sk}$ now becomes

$$f(u_{s1}, \ldots, u_{sk}) = \exp\left(\theta_s \sum_{q=1}^{k} a_q u_{sq} - \sum_{q=1}^{k} b_q a_q u_{sq}\right) \Bigg/ \prod_{q=1}^{k}\left(1+e^{(\theta_s-b_q)a_q}\right) . \qquad [20]$$

As first noted by Birnbaum (1957) and further discussed in his contribution to Lord and Novick (1968), the sufficient statistic for θ_s in this model is

$$t_s = \sum_{q=1}^{k} a_q u_{sq} \;, \qquad [21]$$

so that the latent individual parameter is scored by multiplying the responses with the item discriminations. This is in contrast to the Rasch model, where the scoring of individuals is independent of the parameters to be estimated. It is possible to obtain ML estimates for the various parameters of this two-parameter logistic model, although there are numerical problems. As yet, little is known about the statistical properties of these estimates. It seems, however, that actual data exhibits a structure with varying item discriminations in many applications. In addition, even rough estimates of the a_q's can shed light on possible modifications of the model.

Goodness-of-Fit Tests

There have been a number of suggestions in the literature for checking a latent trait model. The most direct approach is to calculate the probabilities for all the possible response patterns and then to compare the observed and the expected numbers by means of an ordinary χ^2 test. Such a procedure will be considered in connection with the latent class model. Just recently there have been a number of papers that describe and discuss various alternative goodness-of-fit tests. References can be made to Gustafsson (1980) and van den Wollenberg (in press). In this paper the test for the Rasch model suggested by Andersen (1973) based on CML estimation of the item parameters, will be applied.

Multiplying Equation 16 over all individuals gives the conditional probability of all observed response patterns, given the individual scores. In CML estimation this is considered the likelihood function. It depends only on $(b_1, ..., b_k)$ and the observations. For simplicity $L_c(b_1, ..., b_k)$ is written for this function.

As it only depends on the b's, L_c can be expected to give the same estimates apart from random errors when derived from different samples of individuals. The model can, therefore, be checked by comparing the estimates from various subsamples, if such subsamples are large enough to guarantee reasonable small standard errors on the estimates. In choosing the subsamples the aim should, in addition, be to find groups of individuals with very different θ's, as the test is based on examining whether the conditional likelihood really is θ-independent. The individual score $u_{s.}$ is a criterion that is likely to discriminate between individuals with high, medium, and low values of θ. Hence, the test is based on estimating $b_1, \ldots, b_k$ from various score groups.

Define, therefore, R such score groups ranging from low to high scores and denote by $L_c^{(r)}(b_1, \ldots, b_k)$ the conditional likelihood for the individuals in score group r. If $(\hat{b}_1, \ldots, \hat{b}_k)$ are the overall CML estimates and $(\hat{b}_1^{(r)}, \ldots, \hat{b}_k^{(r)})$ the CML estimates from score group r, these estimates can be compared either graphically or by a goodness-of-fit test, which takes the form

$$z = 2\left[\sum_r \ln L_c^{(r)}(\hat{b}_1^{(r)}, \ldots, \hat{b}_k^{(r)}) - \ln L_c(\hat{b}_1, \ldots, \hat{b}_k)\right] \qquad [22]$$

As proved in Andersen (1973), this test statistic is approximately χ^2 distributed with $(R-1)(k-1)$ degrees of freedom.

As z is always positive and close to zero if the $\hat{b}_q^{(r)}$'s are close to the overall estimates $\hat{b}_q$, the model is rejected if the observed value of z is larger than, for example, the 95th percentile of the χ^2distribution with $(R-1)(k-1)$ degrees of freedom. A discussion of the power of this test against the alternative of a two-parameter logistic model is given in Andersen (1973).

Illustrative Data

Table 5 shows the score group estimates $\hat{b}_j^{(r)}$ for three score groups of the complaint data. In Figure 1 the score group estimates are plotted against the overall estimates, with the short lines corresponding to twice the standard error of the $\hat{b}_j^{(r)}$'s measured from the identity line. The test statistic z becomes $z = 17.4$, $df = 10$. As $\chi^2_{0.95}(10) = 18.3$, there is a reasonably good fit, and it can be concluded that the Rasch model fit the data well.

Table 5
Overall and Score Group Estimates of the Item Parameters

			Score Group					
Item	Overall	SE	0-2	SE	3	SE	4-6	SE
1	-2.47	.21	-2.64	.40	-2.36	.33	-2.47	.42
2	-.64	.11	-.69	.29	-.69	.18	-.61	.19
3	-.48	.11	.02	.33	-.79	.18	-.42	.18
4	.66	.10	.02	.33	.44	.17	.76	.14
5	.96	.10	.83	.42	1.36	.22	.81	.14
6	1.97	.10	2.46	.84	2.04	.28	1.93	.13

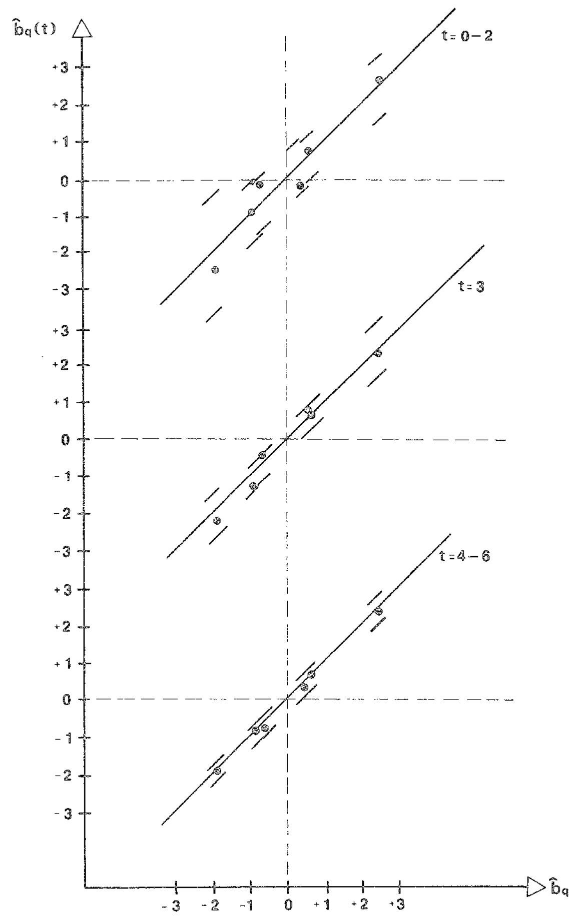
An evaluation of the fit of the model can also be based directly on comparing the observed and expected response patterns under the model. There will not be an assumption of any knowledge of the form of the latent density $\varphi(\theta)$. Hence, the probability

$$\varphi_r = P(U_{s.} = r) \qquad [23]$$

is estimated by the observed frequency n_r/n of individuals in score group r. In this way the probability of response pattern $u_1, \ldots, u_k$ is

$$f(u_1, \ldots, u_k) = f(u_1, \ldots, u_k | r)(n_r/n) \qquad [24]$$

Figure 1
Score Group Estimates of Item Parameters
Plotted Against Overall Parameters



and the estimated expected numbers are

$$n\,\hat{f}(u_1, \ldots, u_k) = n_r[\hat{f}(u_1, \ldots, u_k|r)] \qquad [25]$$

where $\hat{f}(u_1, \ldots, u_k|r)$ is the conditional probability of Equation 16 with the estimates $\hat{b}_1, \ldots, \hat{b}_k$ inserted. In Table 6 the observed and expected response patterns are shown as columns (a) and (b). The two additional columns (c) and (d) refer to models that will be discussed in subsequent sections. As the observed and expected numbers are small for the last 32 response patterns, these are omitted in Table 6.

Latent Class Analysis

In the latent class model of Equation 8, there is for each latent class k probabilities $p_\nu^{(q)}$, $q = 1, \ldots, k$ for a positive answer on item j given latent class ν. As there are M latent classes, this gives

Table 6
Observed Number of Individuals for Selected Response Patterns and Corresponding Expected Numbers Under Three Models

Response Pattern	(a) Observed N_u	(b) Latent Trait n_r/n	(c) Latent Trait $\varphi(\theta\|\mu,\sigma^2)$	(d) Latent Class m = 3
1 1 1 1 1 1	127	127.0	117.9	129.5
1 1 1 1 1 0	78	83.1	92.2	77.5
1 1 1 1 0 1	16	30.2	33.5	16.0
1 1 1 1 0 0	54	50.7	52.3	60.3
1 1 1 0 1 1	36	22.3	24.8	33.5
1 1 1 0 1 0	36	37.4	38.6	35.7
1 1 1 0 0 1	14	13.6	14.0	12.4
1 1 1 0 0 0	59	51.1	40.7	51.0
1 1 0 1 1 1	8	7.2	8.0	7.9
1 1 0 1 1 0	16	12.1	12.5	11.8
1 1 0 1 0 1	4	4.4	4.5	4.7
1 1 0 1 0 0	18	16.5	13.1	19.7
1 1 0 0 1 1	1	3.2	3.3	3.2
1 1 0 0 1 0	6	12.2	9.7	8.7
1 1 0 0 0 1	0	4.4	3.5	4.0
1 1 0 0 0 0	18	15.9	18.5	16.8
1 0 1 1 1 1	12	6.1	6.8	9.4
1 0 1 1 1 0	7	10.3	10.6	10.5
1 0 1 1 0 1	4	3.7	3.8	3.7
1 0 1 1 0 0	18	14.0	11.1	15.4
1 0 1 0 1 1	1	2.8	2.8	3.2
1 0 1 0 1 0	6	10.3	8.2	7.1
1 0 1 0 0 1	4	3.8	3.0	3.1
1 0 1 0 0 0	9	13.5	15.7	13.1
1 0 0 1 1 1	1	.9	.9	.9
1 0 0 1 1 0	4	3.3	2.7	2.6
1 0 0 1 0 1	1	1.2	1.0	1.2
1 0 0 1 0 0	10	4.3	5.1	5.1
1 0 0 0 1 1	1	.9	.7	.6
1 0 0 0 1 0	3	3.2	3.7	2.1
1 0 0 0 0 1	1	1.2	1.4	1.0
1 0 0 0 0 0	3	10.1	12.8	4.3

$M \times k$ $p_v^{(q)}$'s and $M - 1$ φ_v's. For the binary case this compares to the $2^k - 1$ free item response probabilities. Thus, the model is only identifiable if $2^k > M(k+1)$. For the complaint data, where $k = 6$ and $2^k = 64$, models with up to nine latent classes can thus be fitted, but ordinarily the number of latent classes will be kept relatively small to facilitate an interpretation.

Establishing the equations to be solved in order to obtain ML estimates for the parameters is straightforward. For many years it was considered rather difficult to solve the likelihood equations, but this has been a misconception. Formann (1978) described a simple computer program and Goodman (1979) showed that the estimates could be obtained by a simple marginal fitting algorithm, now known as the EM algorithm, which was introduced by Dempster, Laird, and Rubin (1977). Details of the formulas will not be given here. Explicit expressions are available in Goodman (1978, chaps. 8–10) or Haberman (1978, chaps. 10). Instead, it will be shown how the method worked on the complaint data. The calculations are all from Poulsen (1981), who used his own computer program.

Illustrative Data

In Table 7 the parameter estimates for a model with three latent classes are shown. Poulsen describes the three classes as "complainers," "noncomplainers," and "scalables." These descriptive terms refer, of course, to the behavior of the $p_\nu^{(q)}$'s, where there is generally a very high probability of complaining in Class 3, a generally very small probability of complaining in Class 2, and a probability of complaining dependent on the item, i.e., the consumer situation, in Class 1. The relatively high percentage in Class 3 corresponds to the high frequencies of response patterns with many 1's.

Table 7
Estimates of the Parameters $\hat{p}_\nu^{(q)}$ of a 3-Class Latent Model for the Complaint Data

Item	Latent Class (ν) 1 $\hat{\varphi}_\nu=.59$	2 $\hat{\varphi}_\nu=.02$	3 $\hat{\varphi}_\nu=.39$
1	.96	.00	1.00
2	.80	.12	.94
3	.75	.08	.96
4	.54	.00	.82
5	.33	.12	.98
6	.19	.00	.72

In column (d) of Table 6, the estimates of the expected number of individuals for each response pattern are shown. As can be seen, the model gives a relatively good fit to the observed numbers. By collapsing a number of patterns the model may be checked by a χ^2 test. Based on 33 groups of response patterns, this gives a value of $z = 23.6$, $df = 12$, which is between the 95th and 97.5th percentiles of a χ^2 distribution, indicating a less than satisfactory fit.

It is now interesting to compare this type of analysis with the analysis by the latent trait Rasch model. As there are two relatively good fits, it should be expected that the two models will be more or less the same model. One way to illustrate that this is in fact the case, is to check whether the three latent classes should correspond to three intervals on the latent scale of θ. As φ_ν is the probability of $\theta_{\nu-1} < \theta < \theta_\nu$, $p_\nu^{(q)}$ can be interpreted as an average value of $p^{(q)}(\theta)$ in this interval. Since the exact form of $p^{(q)}(\theta)$ under the Rasch model is known, the equation

$$e^{(\theta-b_q)} \Big/ \left(1 + e^{(\theta-b_q)}\right) = p_\nu^{(q)} \qquad [26]$$

can be solved for each value of q. This should give solutions for θ in the interval $(\theta_{\nu-1}, \theta_\nu)$. Since individuals in the ν^{th} latent class are actually assumed to have identical latent values θ, it should even be expected that the q solutions to Equation 26 are very close.

In Table 8, the solutions to Equation 26 are shown for all three latent classes and $q = 1, ..., k$. Table 8 clearly indicates that the two models are close. Since the θ's do not vary much within a latent class and are clearly separated between classes, the $p_\nu^{(q)}$'s can be modeled by a logistic latent trait model. Table 8, on the other hand, also clearly indicates that an assumption of strictly equal latent value within a class is very doubtful. Finally, the latent class model represents a clear overparametrization, as there are 20 parameters, while for the Rasch model there are only 11, namely, 5 item parameters and 6 marginal score frequencies.

Table 8
Solutions to Equation 26 for a Model
with Three Latent Classes

	Latent Class		
Item	Scalables	Non-Complainers	Complainers
1	.7	-7.1	2.1
2	.7	-2.6	2.1
3	.6	-2.9	2.7
4	.8	-3.9	2.2
5	.2	-1.0	4.9
6	.5	-2.6	2.9

A slightly different way to compare results from the two types of analysis is to determine to which latent class each response vector most likely belongs, and to compare this with the score group. An individual is assigned to latent class ν, if for his/her observed response pattern $u_1, ..., u_k$ the probability

$$\prod_{q=1}^{k} (p_\nu^{(q)})^{u_q} (1-p_\nu^{(q)})^{1-u_q} \varphi_\nu \qquad [27]$$

is larger than the corresponding probabilities for the other classes. In Table 9 the observed number of individuals for each combination of assigned latent class and observed raw score are shown. As can be seen, only 21 individuals, or 3.5%, will be differently classified if Classes 1, 2, and 3 correspond to Scores 2-4, 0-1 and 5-6.

When the degree of fit between columns (b) and (d) of Table 6 is compared, of course there is a better fit by the latent class model with almost twice as many parameters. When the goodness-of-fit test previously given for the latent class model is compared with a corresponding test obtained from column (b) in Table 6, where the same response patterns have been collapsed as for column (d), this gives $z = 48.6$, $df = 22$. There is actually not a very much better fit, considering the larger number of degrees of freedom in this last test.

Table 9
Joint Distribution of Assigned Latent Class and Observed Raw Score for Complaint Data

Assigned Latent Class	Score Group 0	1	2	3	4	5	6
1	0	3	42	124	141	16	0
2	7	4	1	0	0	0	0
3	0	0	0	0	1	134	127

Latent Structure Analysis

This section will show some results, assuming a continuous latent density $\varphi(\theta)$. Thus, it is necessary to go back to Equation 11 and assume that $\varphi(\theta|\mu, \sigma^2)$ is a normal distribution density with mean value μ and variance σ^2. As the conditional distribution of $u_{s1}, \ldots, u_{sk}$ given $u_{s.} = r$ (i.e., the response pattern given the score) is independent of θ_s, it is only necessary to consider the probability distribution of the scores in order to make inferences concerning the density $\varphi(\theta|\mu, \sigma^2)$.

From Equation 14 is obtained the distribution of $u_{s.} = r$ given θ_s as

$$f(r|\theta_s) = e^{\theta_s r} G_r(b_1, \ldots, b_k) \Big/ \prod_{q=1}^{k} \left(1 + e^{(\theta_s - b_q)}\right) . \quad [28]$$

and, hence, the marginal distribution of r as

$$f(r) = G_r(b_1, \ldots, b_k) \int \frac{e^{\theta r} \varphi(\theta|\mu,\sigma^2)}{\prod_q \left(1 + e^{(\theta - b_q)}\right)} d\theta \quad [29]$$

where $G_r(b_1, \ldots, b_k)$ is a certain function of $b_1, \ldots, b_k$. It follows that the distribution of the observed score group counts $n_0, \ldots, n_k$ is multinomial with cell probabilities $\pi_0, \ldots, \pi_k$, where

$$\pi_r = \pi_r(\mu,\sigma^2) = f(r) . \quad [30]$$

For given values of $(b_1, \ldots, b_k)$, ML estimates $\hat{\mu}$ and $\hat{\sigma}^2$ of μ and σ^2 can then be obtained and the model checked by

$$z = 2 \sum_{r=0}^{k} n_r \{\ln(n_r) - \ln[n\pi_r(\mu,\sigma^2)]\} \quad [31]$$

which (still for given values of $b_1, \ldots, b_k$) is approximately χ^2-distributed with $df = k - 2$ degrees of freedom, as the multinomial distribution has $k + 1$ cells and two parameters are estimated (cf. Andersen & Madsen, 1977).

If the b_q's are not known, the estimates $\hat{b}_1, \ldots, \hat{b}_k$ can be used. The test statistic is now not exactly χ^2-distributed, but for relatively precise estimates, the percentiles of the χ^2-distribution with $df = k - 2$ will still give an indication of the goodness of fit of the model to the data.

Illustrative Data

The complaint data of Table 2 result in the estimates $\hat{\mu} = 1.40$ and $\hat{\sigma}^2 = 1.42$. When inserted in Equation 31 (score groups $r = 0$ and 1 are collapsed), this gives a goodness-of-fit test statistic of $z = 12.0$, $df = 3$, so that a latent normal density is clearly rejected.

The expected and observed values are shown in Table 10. From Table 10 it seems that the observed frequencies exhibit a flatter curve at its maximum than can be accounted for by the normal latent density.

Table 10
Observed and Expected Numbers in Each Score Group With a Normal Latent Density

Score Group	Observed	Expected
0	7	2.6
1	7	17.8
2	43	50.2
3	124	98.6
4	142	146.5
5	150	166.4
6	127	117.9
Total	600	600

In column (c) of Table 6 the expected response patterns under the Rasch model with a normal latent density are shown. The poorer fit as compared with the latent class model and with the Rasch model and a complete fit of score group numbers is very obvious. A χ^2 test parallel to those earlier computed from columns (b) and (d) of Table 8 gives a value of $z = 70.1$, $df = 25$, which is highly significant.

The analysis described above has been further discussed in Andersen and Madsen (1977), Andersen (1980a), and Sanathanan and Blumenthal (1978). Similar models have been discussed by Muthén (1978, 1979), Christofferson (1975), Bock (1972), Bartholomew (1980), and Bock and Aitkin (1981). A recent review of various models and the connection to contingency tables has been given by Andersen (1982). Tjur (1982) has discussed the relationship between the tests to be applied at various stages of the model control.

In several of these papers ML estimation is not applied. As the likelihood equations are somewhat difficult to handle, various forms of weighted least squares are used. It often turns out that simple computer algorithms give satisfactory estimates for use in the model check tests.

There is, of course, the problem of choosing a latent density. Often several possible densities will give approximately equally good fit, and in such situations it may be desirable to choose a density with easily interpretable parameters. Considerations of this sort are behind the choice of the normal density above. When a normal density does not fit very well, alternatives may be explored. For the case of the complaint data, however, this has as yet not been done. In Andersen (1980a) there are several suggestions for extending the method to cover comparisons of several populations and time-dependent models.

References

Andersen, E. B. A goodness of fit test for the Rasch model. *Psychometrika*, 1973, *38*, 123-140.

Andersen, E. B. Comparing latent distributions. *Psychometrika*, 1980, *45*, 121-134. (a)

Andersen, E. B. *Discrete statistical models with social science applications*. Amsterdam: North Holland Publishing Co., 1980. (b)

Andersen, E. B. Latent structure analysis. A review. *Scandinavian Journal of Statistics*, 1982, *9*, 1-12.

Andersen, E. B., & Madsen, M. Estimating the parameters of the latent population distribution. *Psychometrika*, 1977, *42*, 357-374.

Baker, F. B. Advances in item analysis. *Review of Educational Research*, 1977, *47*, 151-178.

Barndorff-Nielsen, O. *Information and exponential families in statistical theory*. New York: John Wiley, 1978.

Bartholomew, D. J. Factor analysis for categorical data. *Journal of the Royal Statistical Society, Series B*, 1980, *42*, 293-321.

Birnbaum, A. *Efficient design and use of tests of a mental ability for various decision-making problems*. (Series Report No. 58-16, Project No. 7755-23). Randolph Air Force Base TX: USAF School of Aviation Medicine, January 1957.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. *Discrete multivariate analysis. Theory and practice*. Cambridge: MIT Press, 1975.

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

Bock, R. D. *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill, 1975.

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

Christofferson, A. Factor analysis of dichotomized variables. *Psychometrika*, 1975, *40*, 5-32.

Cohen, L. Approximate methods for parameter estimates in the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 1979, *32*, 113-120.

Dempster, A. P., Laird, N. N., & Rubin, D. B. Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society, Series B*, 1977, *39*, 1-22.

Formann, A. K. A note on parameter estimation for Lazarsfeld's latent class analysis. *Psychometrika*, 1978, *43*, 123-126.

Goodman, L. A. *Analysing qualitative/categorical data. Log-linear models and latent structure analysis*. London: Addison & Wesley, 1978.

Goodman, L. A. On the estimation of parameters in latent structure analysis. *Psychometrika*, 1979, *44*, 123-128.

Gustafsson, J.-E. Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical Statistical Psychology*, 1980, *33*, 205-233.

Haberman, S. J. Maximum likelihood estimation in exponential response models. *Annals of Statistics*, 1977, *5*, 815-841.

Haberman, S. J. *Analysis of qualitative data*. (Vols. 1 & 2). New York: Academic Press, 1978.

Hambleton, R. K., & Cook, L. L. Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement*, 1977, *14*, 75-96.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eigner, D. R., & Gifford, J. A. Developments in latent trait theory: Models, technical issues and applications. *Review of Educational Research*, 1978, *48*, 467-510.

Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 1968, *28*, 989-1020.

Lord, F. M. *Applications of item response theory to practical testing*. Hillsdale NJ: Erlbaum, 1981.

Lord, F. M., & Novick, M. R. *Statistical theories of mental test scores*. Reading MA: Addison & Wesley, 1968.

Muthén, B. Contributions to factor analysis of dichotomous variables. *Psychometrika*, 1978, *43*, 551-560.

Muthén, B. A structural probit model with latent variables. *Journal American Statistical Association*, 1979, *74*, 807-811.

Poulsen, C. S. Latent class analysis of consumer complaining behaviour. In Höskuldsson et al. (Ed.), *Symposium in Applied Statistics*. Copenhagen: NEUCC, 1981.

Rasch, G. *Probalistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press, 1980. (Originally published, Copenhagen: Danmarks Paedagogiske Institut, 1960)

Sanathanan, L., & Blumenthal, S. The logistic model and estimation of latent structure. *Journal of the American Statistical Association*, 1978, *73*, 794-799.

Tjur, T. A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics*, 1982, *9*, 23-30.

Wainer, H., Morgan, A., & Gustafsson, J.-E. A review of estimation procedures for the Rasch model with an eye toward longish tests. *Journal of Educational Statistics*, 1980, *5*, 35-64.

Weinreich, M. *An SAS-implementation of the Rasch item analysis with two categories.* University of Copenhagen: Institute of Statistics, 1980.

Weiss, D. J., & Davison, M. L. Review of test theory and methods. *Annual Review of Psychology,* 1981, *32*, 629-658.

Wright, B. D., & Douglas, G. A. Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement,* 1977, *37*, 573-586.

Wright, B. D., & Stone, M. H. *Best test design.* Chicago: Mesa Press, 1979.

Wollenberg, A. L. Van den. Two new test statistics for the Rasch model. *Psychometrika,* in press.

Author's Address

Send requests for reprints or further information to Erling B. Andersen, Institute of Statistics, University of Copenhagen, Studiestraede 6, DK-1455 Copenhagen K, Denmark.