Tue Tjur

A Connection Between Rasch's Item Analysis Model and a Multiplicative Poisson Model

Preprint
September
1980

6

Institute of Mathematical Statistics
University of Copenhagen

Tue Tjur

A CONNECTION BETWEEN RASCH'S ITEM ANALYSIS MODEL AND A MULTIPLICATIVE POISSON MODEL

Preprint No. 6

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

September 1980

Summary: Introduction of random row effects in Rasch's item analysis model (or the two-way logit-additive model for binary data), followed by a (not very controversial) extension of this model leads to a model which is shown to be equivalent to a conditional multiplicative Poisson model. This result gives som new insight in the structure of Rasch's model and suggests, in particular, some goodness of fit tests which are applicable also to the original "fixed effects" Rasch model.

Key words: item analysis, latent structure, logit-linear model, multiplicative Poisson model, Rasch model.

1. The Rasch model.

Suppose that each of n *subjects* (typically persons), labelled $i = 1,\ldots, n$, are exposed to k *items* (typically questions to be answered, problems to be solved etc.), labelled $j = 1,\ldots, k$. For each item and each subject, a binary response $x_{ij} = 1$ or 0 (e.g. yes or no, correct or incorrect etc.) is recorded. Thus, a two-way table (table 1) is obtained, with

subject \ item	1	2	...	k	
1	1	0		0	r_1
2	0	1		1	r_2
⋮			x_{ij}		
n	0	0		0	r_n
	s_1	s_2		s_k	s

Tabel 1

row marginals $r_1,\ldots, r_n$, the socalled *raw scores*, column marginals $s_1,\ldots, s_k$ and total sum $s = s_1 +\ldots+ s_k = r_1 +\ldots+ r_n$. By n_r $(r = 0,1,\ldots,k)$ we denote the number of subjects in the r'th *score group*, i.e. the number of subjects with raw score r. Thus, we have $n = n_0 + n_1 +\ldots+ n_k$ and $s = 0\, n_0 + 1\, n_1 +\ldots+ k\, n_k$.

Rasch's item analysis model (see e.g. Rasch 1960, 1961) assumes that the x_{ij}'s are independent random variables with

$$p_{ij} = P(x_{ij} = 1) = \frac{\xi_i \lambda_j}{(1 + \xi_i \lambda_j)} \quad ,$$

where $\xi_i > 0$ is the *subject parameter* (expressing, e.g., person no. i's ability in solving problems of the type considered) and $\lambda_j > 0$ is the item parameter (expressing, e.g., the difficulty of problem j, small values of λ_j corresponding to a high degree of difficulty). The main idea is that the probability p_{ij} of a positive response is a function of a product (or, equivalenty, a sum) of an item parameter and a subject parameter. This function is given by the above equation, or, equivalently, by

$$\text{logit } p_{ij} = \log \xi_i + \log \lambda_j.$$

The last equation shows that the model is a linear logistic model (see Cox, 1970), with a linear structure similar to the structure of additivity in an ordinary two-way ANOVA-model.

The model is symmetric in items and subjects (and also in the responses 0 and 1, in fact), but we are mainly interested in the situation where n is large compared with k. Emphasis will be put on goodness of fit tests and estimation of the item parameters, while the subject parameters, are regarded as nuisance. The relevant asymptotic assumption in this situation is $n \to \infty$. Standard asymptotic theory for maximum-likelihood fails in this case, because the number of nuisance parameters increases as $n \to \infty$. It is known that the maximum likelihood estimates of the item parameters are inconsistent in this case, even under the nicest possible assumptions about the variation of the subject parameters (see e.g. Andersen, 1980). However, this problem has a nice solution (see Rasch 1960, 1961, Andersen 1973a, 1980):

Consider the row sums $r_1, \ldots, r_n$. Obviously, these numbers contain very little information about the differences between specific items. Consequently, we loose very little by conditioning on these row sums (an idea which can be supported by more general inference principles, see Andersen, 1973a), and the subject parameters do not occur in the conditional model. To see this, let A_i denote the (random) set of items responded to by subject i, i.e.

$$A_i = \{j \mid x_{ij} = 1\} .$$

For a fixed subset A of $\{1,\dots,k\}$, the probability that subject i obtains exactly this pattern of responses is given by

$$P(A_i = A) = \Big(\prod_{j\in A} p_{ij}\Big)\Big(\prod_{j\in A^c} (1 - p_{ij})\Big)$$

$$= \Big(\prod_{j\in A} \frac{\xi_i\lambda_j}{1+\xi_i\lambda_j}\Big)\Big(\prod_{j\in A^c} \frac{1}{1+\xi_i\lambda_j}\Big)$$

$$= \frac{\xi_i^{\,r} \prod_{j\in A} \lambda_j}{\prod_{j=1}^{k} (1+\xi_i\lambda_j)} ,$$

where $r = \# A$ = the number of elements in A. Conditionally on the row sum $r_i = \# A_i$, the probability of this event is

$$P(A_i = A \mid r_i = r) = \frac{P(A_i = A)}{\sum_{B:\#B=r} P(A_i = B)} =$$

$$\frac{\xi_i^{\,r}\big(\prod_{j\in A} \lambda_j\big) \Big/ \prod_{j=1}^{k} (1 + \xi_i\lambda_j)}{\sum_{B:\#B=r} \xi_i^{\,r}\big(\prod_{j\in B} \lambda_j\big) \Big/ \prod_{j=1}^{k} (1 + \xi_i\lambda_j)} =$$

$$\frac{\prod_{j\in A} \lambda_j}{\sum_{B:\#B=r} \prod_{j\in B} \lambda_j} = \frac{\prod_{j\in A} \lambda_j}{\gamma_r(\lambda_1,\dots,\lambda_k)}$$

where $\gamma_r(\lambda_1,\ldots,\lambda_k) = \sum_{B:\#B=r} \prod_{j\in B} \lambda_j$ (this short notation is used throughout the paper).

This is the probability that a subject i obtains response pattern A when its raw score is known to be r, and the point is that this probability does not depend on the subject parameter ξ_i. The likelihood function L_c (c for conditional) of the conditional model, given $r_1,\ldots,r_n$, is obtained by multiplication over i:

$$L_c(\lambda_1,\ldots,\lambda_k) = \prod_{i=1}^{n} \frac{\prod_{j\in A_i} \lambda_j}{\gamma_{r_i}(\lambda_1,\ldots,\lambda_k)}$$

$$= \frac{\lambda_1^{s_1}\cdots\lambda_k^{s_k}}{\prod_{r=0}^{k} \gamma_r(\lambda_1,\ldots,\lambda_k)^{n_r}} .$$

It requires only slightly more than standard asymptotic theory to see that the maximum-likelihood estimates obtained from this conditional likelihood are <u>conditionally</u> consistent as $n \to \infty$, under the obvious assumption that the increase of n is not only due to increase of the trivial score groups corresponding to $r = 0$ and $r = k$. In fact, the estimates based exclusively on scoregroup r, $1 \leq r \leq k-1$, are conditionally consistent by standard asymptotic theory, provided that $n_r \to \infty$. Consistency (and asymptotic normality) of the conditional estimates holds also with respect to the original (unconditioned) distribution, under suitable restrictions on the behaviour of the increasing set of subject parameters, see Andersen (1973a).

2. The Rasch model with random row effects.

In many applications of the Rasch model it is reasonable to think of the set of subjects as a random sample from a larger population. It seems permissible, then, to introduce an underlying common distribution of the subject parameters, thus expressing the fact that any information about specific subject parameters is regarded as nuisance. This idea is known from latent structure analysis, see e.g. Andersen (1980). Thus, let us assume that $\xi_1, \ldots, \xi_n$ are independent, identically distributed random variables from some (completely unknown) distribution π on the positive half line. In this model, the rows of table 1 are independent, identically distributed, but x_{ij}'s in the same row are no longer independent. Let $q_0, q_1, \ldots, q_k$ denote the point probabilities in the distribution of an arbitrary row sum, i.e.

$$q_r = q_r(\lambda_1, \ldots, \lambda_k, \pi) = P(r_i = r) = \sum_{A:\#A=r} P(A_i = A) = E \sum_{A:\#A=r} P(A_i = A \mid \xi_i)$$

$$= \int \sum_{A:\#A=r} \frac{\xi^r \prod_{j \in A} \lambda_j}{\prod_j (1+\xi \lambda_j)} \pi(d\xi) = \gamma_r(\lambda_1, \ldots, \lambda_k) \int \frac{\xi^r}{\prod_j (1+\xi\lambda_j)} \pi(d\xi) .$$

For any response pattern $A \subseteq 1, \ldots, k$ with $r = \# A$ we have then

$$P(A_i = A) = q_r P(A_i = A \mid r_i = r) .$$

Now, according to the previous section, the conditional probability $P(A_i = A \mid r_i = r)$ is independent of ξ_i in the model with fixed subject parameters. Mixing with respect to the distribution π of ξ_i does obviously not affect this conditional probability, which means that the formula

$$P(A_i = A \mid r_i = r) = (\prod_{j \in A} \lambda_j) / \gamma_r(\lambda_1,\dots,\lambda_k)$$

holds also in the model with random row effects. Thus, the unconditional distribution of the i'th row is given by

$$P(A_i = A) = q_r(\prod_{j \in A} \lambda_j) / \gamma_r(\lambda_1,\dots,\lambda_k) \quad (r = \# A),$$

and the likelihood function L_r (r for random) becomes

$$L_r(\lambda_1,\dots,\lambda_k, \pi) = \prod_{i=1}^{n} \frac{q_{r_i} \prod_{j \in A_i} \lambda_j}{\gamma_{r_i}(\lambda_1,\dots,\lambda_k)}$$

$$= \frac{\lambda_1^{s_1} \dots \lambda_k^{s_k}}{\prod_{r=0}^{k} \gamma_r(\lambda_1,\dots,\lambda_k)^{n_r}} \, q_0^{n_0} q_1^{n_1} \dots q_k^{n_k} .$$

As we have seen, the probabilities $q_0,\dots, q_k$ are complicated functions of the unknown parameters $\lambda_1,\dots,\lambda_k$ and π, and an attempt to maximize the likelihood directly as a function of $(\lambda_1,\dots,\lambda_k,\pi)$ would hardly be successful. However, the form of the likelihood function shows that our model is a submodel of a more attractive model, namely the one obtained by allowing the vector $(q_0, q_1,\dots, q_k)$ to vary freely on the k-dimensional probability simplex.

This "extended random model" (as we shall call it) is thus given by a likelihood function equal to the expression above, but now regarded as a function L_e (e for extended) of the λ_j's and the q_r's , $\lambda_j > 0$, $q_r \geq 0$, $\Sigma_0^k q_r = 1$.

Notice that we have

$$L_e(\lambda_1,\ldots,\lambda_k, q_0, q_1,\ldots, q_k) =$$

$$L_c(\lambda_1,\ldots,\lambda_k)\; q_0^{n_0}\; q_1^{n_1} \ldots q_k^{n_k} \;,$$

in agreement with the fact that the extended random model has an alternative interpretation as a randomized version of the <u>conditional</u> Rasch model. Suppose, namely, that the whole experiment can be simulated as follows: First, the row sums $r_1,\ldots,r_n$ are generated as random numbers with distribution $P(r_i = r) = q_r$. Then, the contents of the rows (with given row sums) are generated according to the conditional distribution found in section 1. This would obviously lead to the likelihood function L_e above.

While the Rasch model with random row effects is a usual statistical construction, it is, perhaps, less obvious what the extended random model really states about data. Martin-Löf (1970) has discussed this model and characterized it as the exponential family in which the column sums and the empirical distribution of the row sums (i.e. $s_1,\ldots,s_k$ and $n_0, n_1,\ldots,n_k$) are sufficient. An analogy with a simpler and more familiar model may help to settle the ideas:

Consider an ordinary two-way ANOVA-model, i.e.

$$x_{ij} \sim N(\xi_i + \lambda_j, \sigma^2) \;,\; i = 1,\ldots,n \;,\; j = 1,\ldots,k \;.$$

Introduction of a random row effect in this model can be carried out in two different ways, namely

1) by introduction of an unknown normal distribution, describing the variation of the row parameters ξ_i, and

2) by introduction an unknown normal distribution, describing the variation of the row sums $r_i = \sum_j x_{ij}$, the remaining random variation being described by the original conditional distributions, given the row sums, which (also in this case) are independent of the parameters ξ_i.

It is wellknown that the second model is an extension of the first one, namely the extension obtained by allowing for negative correlations in a two-way ANOVA-model with random row effects. Mathematically, the second model is the more attractive, but a negative correlation may be diffecult to explain in certain applied situations.

3. A multiplicative Poisson model.

Consider the following alternative way of representing data like those given by table 1: Each item $j = 1,\dots,k$ is regarded as a classifying factor according to which subjects are classified into two groups. Cross-classifying the n subjects according to these k factors, we obtain a 2^k- contingency table, the cells of which we may label by subsets A of $1,\dots,k$. The count n_A of the A'th cell is simply the number of subjects with response pattern A.

The data reduction $(x_{ij}) \to (n_A)$ is obviously a sufficient transformation in the extended random model of the previous section. Hence, it is possible to consider this as a model for the counts n_A, without really changing anything. But it is more tempting to search among models related to those usually applied to contingency tables, namely the multiplicative Poisson models and the models derived from these by conditioning on suitable marginals.

However, the multiplicative Poisson models usually applied to contingency tabels are not relevant. The simplest model (the model of "independence"), stating that the counts n_A are independent, Poisson distributed with

$$E\, n_A = \mu \prod_{j \in A} \lambda_j$$

is only relevant if the population of subjects can be assumed homogeneous. Indeed, independence of the responses of a random subject to different items can only be assumed when the subject are known to be equally "clever", otherwise we would certainly expect a positive correlation. This is also true in the more precise sense that (n_A) is a sufficient reduction in the Rasch model with a common subject parameter $\xi = \xi_1 = \ldots = \xi_n$, and this reduction of data leads to the model obtained by conditioning on $n = \Sigma\, n_A$ in the simple multiplicative Poisson model above.

Thus, what we need is an extension of this model, taking "ability" of subjects into account. An explicit introduction of higher order interactions would not be easy to justify from this point of view. A more constructive idea is to extend the model by an additional factor, classifying subjects according to "ability", as far as this property of subjects is reflected by data. The obvious candidate for such a factor is the classification into raw score groups. This leads to the multiplicative model obtained by letting the constant μ in the independence model above depend on the score group, i.e.

$$E\, n_A = \mu_{\#A} \prod_{j \in A} \lambda_j \; .$$

This is the model which we shall study in the following.

The model has 2k+1 parameters, namely the item parameters $\lambda_1,\ldots,\lambda_k$ and the "raw score parameters" $\mu_0, \mu_1,\ldots,\mu_k$. However, the model is overparameterized, since replacement of λ_j by λ_j/c and μ_r by $\mu_r\, c^r$, $c > 0$, leaves the distribution unchanged. Apart from this, no overparametrization occurs, which means that the intrinsic dimension of the model is 2k. The overparametrization can be avoided by assuming $\Pi\, \lambda_j = 1$ or $\mu_k = 1$, but there will be no need for such normalizations in the present paper.

Our main observation is this: <u>Apart from the randomness imposed on the total number n of subjects, this multiplicative Poisson model is equivalent to the extended random model of section 2</u>.

In order to prove this, consider the following transformation of the likelihood function L_p (p for Poisson):

$$L_p(\lambda_1,\ldots,\lambda_k\, ,\, \mu_0\, ,\, \mu_1,\ldots,\mu_k) =$$

$$\prod_{A \subseteq \{1,\ldots,k\}} \exp(-\mu_{\#A} \prod_{j \in A} \lambda_j)\; \frac{(\mu_{\#A} \prod_{j \in A} \lambda_j)^{n_A}}{n_A!} =$$

$$\left[\frac{n!}{\prod_A n_A!}\right] \left[\exp(-\sum_A \mu_{\#A} \prod_{j \in A} \lambda_j)\; \frac{(\sum_A \mu_{\#A} \prod_{j \in A} \lambda_j)^n}{n!}\right]$$

$$\left[\prod_{r=0}^{k} \left(\frac{\mu_r \gamma_r(\lambda_1,\ldots,\lambda_k)}{\sum_{r'=0}^{k} \mu_{r'} \gamma_{r'}(\lambda_1,\ldots,\lambda_k)}\right)^{n_r}\right]$$

$$\left[\frac{\prod_{A} \left(\prod_{j \in A} \lambda_j \right)^{n_A}}{\prod_{r=0}^{k} \gamma_r(\lambda_1, \ldots, \lambda_k)^{n_r}} \right] .$$

The formal proof of this identity is straightforward, when it is noticed that $\sum_{A} \mu_{\#A} \prod_{j \in A} \lambda_j = \sum_{r'=0}^{k} \mu_{r'} \gamma_{r'} (\lambda_1, \ldots, \lambda_k)$. The intuitive content of the formula is this: Apart from combinatorial constants (which are collected in the first pair of brackets), the right hand side is simply the way we would write the likelihood function if the experiment was carried out by first observing n (the second pair of brackets), then observing the numbers n_r of subjects in the score groups in their conditional distribution given n (the third pair of brackets), and finally observing the distribution of response patterns within score groups, given the n_r's (the last pair of brackets).

Now, define

$\lambda = \sum_{A} \mu_{\#A} \prod_{j \in A} \lambda_j$ (= the parameter of the Poisson distribution of n)

and

$q_r = \frac{1}{\lambda} \mu_r \gamma_r (\lambda_1, \ldots, \lambda_k)$ (= the probability that a given subject falls in scoregroup r when n is given).

If we parameterize the model by $\lambda_1, \ldots, \lambda_k, \lambda, q_0, q_1, \ldots, q_k$ instead of $\lambda_1, \ldots, \lambda_k, \mu_0, \mu_1, \ldots, \mu_k$, the likelihood becomes

$$\left[\frac{n!}{\prod_{A} n_A !} \right] \left[e^{-\lambda} \frac{\lambda^n}{n!} \right] \left[q_0^{n_0} q_1^{n_1} \ldots q_k^{n_k} \right] \left[\frac{\lambda_1^{s_1} \ldots \lambda_k^{s_k}}{\prod_{r=0}^{k} \gamma_r(\lambda_1, \ldots, \lambda_k)^{n_r}} \right] =$$

$$\left[\frac{n!}{\prod_A n_A!}\right]\left[e^{-\lambda}\frac{\lambda^n}{n!}\right]\left[q_0^{\;n_0}\, q_1^{\;n_1}\cdots q_k^{\;n_k}\right] L_c(\lambda_1,\ldots,\lambda_k) =$$

$$\left[\frac{n!}{\prod_A n_A!}\right]\left[e^{-\lambda}\frac{\lambda^n}{n!}\right] L_e(\lambda_1,\ldots,\lambda_k\,,\,q_0\,,\,q_1,\ldots,q_k)\ .$$

Thus, apart from the factor $e^{-\lambda}\,\lambda^n/n!$, which can obviously be cancelled by conditioning on n, and a combinatorial factor reflecting the fact that only <u>counts</u> of subject with given response patterns are described by the contingency table, we have the same likelihood as in the extended random model of the previous section. The reparametrization $(\lambda_1,\ldots,\lambda_k\,,\,\mu_0\,,\,\mu_1,\ldots,\mu_k) \to (\lambda_1,\ldots,\lambda_k,$ $q_0\,,\,q_1,\ldots,q_k\,,\,\lambda)$ is one-to-one, and the induced domain of variation for the new parameters is given by the obvious conditions $\lambda_j > 0\,,\, q_r \geq 0\,,\,\lambda > 0\,,\, q_0 + q_1 + \ldots + q_k = 1$. Indeed, for $\lambda_1,\ldots,\lambda_k$ fixed it is easy to see that any $\lambda\,,\,q_0\,,\,q_1,\ldots,q_k$ can be obtained by suitable (unique) choice of $\mu_0\,,\,\mu_1,\ldots,\mu_k$. This proves the desired result.

Notice that the likelihood function L_p decomposes as a product of the function L_c of the item parameters and a function of the remaining parameters. From this it follows immediately that the maximum likelihood estimates of the item parameters (and also the second derivatives of the log likelihood with respect to the item Parameters) in our multiplicative Poisson model coincide with those of the conditional Rasch model. This is computationally convenient, because it means that estimation in the conditional Rasch model can be carried out by means of standard algorithms for estimation in multiplicative Poisson models, like those available in the Rothamsted programming languages GLIM and GENSTAT (described in Nelder and Wedderburn (1972)) and the program FREQ described by Haberman (1979). The more slowly convergent (but computationally easy) algorithm of iterative proportional scaling (see Haberman (1978)) is also applicable, and is, in fact, equivalent to an iterative algorithm for estimation in the conditional Rasch model suggested by Martin-Löf (1970), based on the iteration

$$\lambda_j := s_j / \left(\sum_{r=0}^{k} n_r \frac{\gamma_{r-1}(\lambda_1, \ldots, \lambda_{j-1}, \lambda_{j+1}, \ldots, \lambda_k)}{\gamma_r(\lambda_1, \ldots, \lambda_k)} \right) .$$

Similar remarks apply to the goodness of fit tests suggested in the following sections, since these are stated primarily as likelihood ratio tests in multiplicative Poisson models.

4. Control of the Rasch model, general remarks.

The subsumed idea behind the Rasch model is that the k items represent ways of measuring the <u>same</u> property of subjects. Deviations from the model will typically occur when the response to a single item (or a small group of items) depends on other subject properties than do the responses to the main body of items.

It should be noticed, however, that acceptance of the Rasch model by whatever statistical goodness of fit test can never, in itself, prove the validity of an underlying one-dimensional structure of the type "ablility" of subjects versus "difficulty" of items. An illustrative (imagined) example is this: Suppose that the subjects are school-children, and that the items are simple tests of basic knowledge, e.g. questions like "when was the second world war", "what does a beaver eat", "how does a bear survive the winter", etc. It would usually not be of great interest to analyze such data by means of a Rasch model, mainly because the structure of childrens knowledge is believed to be more interesting than accounted for by a single parameter. In particular, we would certainly not expect a Rasch model to hold for the three items suggested above, because the two last questions are, in an obvious sense, more closely related to each other than to the first question. This will be explained more carefully in section 5. However, for other choices of the questions constituting the item set, we might happen to observe data to which a Rasch model would fit. Roughly speaking, this can happen if the qustions are "equally unrelated", in the sense that no pair of questions are closer related to each other than any other pair. In this case, the subject parameters would, perhaps, be interpretable as "degrees of general knowledge", while all sorts of topic-related knowledge would only contribute to the random variation of the responses. But the extension of the item set by a new question, closely related to a question already occuring in the item set, would tend to destroy the whole thing.

In psychological testing it is usually not possible to apply an item more than once to a subject. The present study of the Rasch model originated in a medical case-study where a number of persons (subjects) were exposed to a number of grass-pollen extracts (items) by the socalled skin prick test. The interpretation of the Rasch model in this case (cfr. similar applications of logit-linear models in bio-assay, see e.g. Finney, 1971) would obviously be that the allergenes behave as if they were different concentrations of a single allergene. In this experiment it would have been possible (unfortunately it was not done) to include different concentrations (or even repetitions of the same concentration) of the same allergene as different items. In such situations - and only in such situations - where some of the items are known to measure the same property of subjects, does a statistical acceptance of the Rasch model really confirm the hypothesis of an underlying one-dimensional logit-additive structure.

5. Control of the multiplicative Poisson model.

The immediate suggestion for a goodness of fit test of the multiplicative Poisson model of section 3 is the likelihood-ratio test of this 2k-dimensional model against the full 2^k-dimensional model specified by a separate Poisson distribution of each count n_A. However, approximate χ^2-distribution (with $f = 2^k - 2k$) of the statistic

$$-2 \log Q = -2 \log \left((\max L_p) / (\prod_A e^{-n_A} n_A^{n_A} / n_A!) \right)$$

requires fitted cell means which are not too small. Since the average of all the counts n_A is $n/2^k$, this approximation is only reliable for small values of k and/or very large values of n.

One way of tackling this problem is to set up a smaller alternative hypothesis somewhere in between our multiplicative Poisson model and the full model. An appropriate choice is a multiplicative Poisson model with a separate set of item parameters for each score group, i.e.

$$E\, n_A = \mu_{\# A} \prod_{j \in A} \lambda_{j,\# A} \quad .$$

The number of parameters in this model is $(k+1) + (k+1)\ k = (k+1)^2$, but the intrinsic dimension of the model is only $(k+1)^2 - (k-1) - 2k = k^2 - k + 2$, because the parameters μ_r for $r \neq 0$, k and $\lambda_{j,r}$ for $r = 0$, k can be set to 1 without changing the model. Thus, the $-2 \log Q$ -statistic for test of our original multiplicative Poisson model against this extended model (which we shall not write down explicitely, since the expression would be rather uninformative) is approximately χ^2 -distributed with $(k^2-k+2) - (2k) = (k-1)(k-2)$ degrees of freedom, and this approximation is obviously less sensitive to small cell counts. This test (or rather, the corresponding conditional test in the Rasch model, see section 6) was suggested by Martin-Löf (1970) and Andersen (1973a, 1973b).

Another way of avoiding the problem of small expected counts is to test the model against the full model, restricting to selected smaller subsets of the item set. It is easy to show (and intuitively not surprising) that validity of our multiplicative Poisson model implies validity of the same model for any subset of the item set (with the same item parameters, but with raw score parameters μ_r depending in a more complicated way of the parameters of the big model). $k = 3$ is the smallest number of items for which a test of the model makes sense (for $k = 2$, the model coincides with the full model), so it may be a good idea to start by looking for deviances from the model restricting to triples of items. For $k = 3$, our multiplicative Poisson model is specified by table 2, which for each $A \subseteq \{1,2,3\}$

	$\emptyset$ μ_0	
{1} $\mu_1 \lambda_1$	{2} $\mu_1 \lambda_2$	{3} $\mu_1 \lambda_3$
{1,3} $\mu_2 \lambda_1 \lambda_3$	{2,3} $\mu_2 \lambda_2 \lambda_3$	{1,2} $\mu_2 \lambda_1 \lambda_2$
	{1,2,3} $\mu_3 \lambda_1 \lambda_2 \lambda_3$	

Table 2

gives the parameter in the Poisson distribution of n_A. The model is 6-dimensional, so the likelihood-ratio statistic $-2 \log Q$ is approximately χ^2 -distributed with $2^3 - 6 = 2$ degrees of freedom.

Looking closer at table 2, we see that this model implies independence in certain 2×2 - subtables. Take, for example, the 2×2 - table consisting of the two first cells of the two middle rows. This is given as table 3, and we notice

{1} $\mu_1 \lambda_1$	{2} $\mu_1 \lambda_2$
{1,3} $\mu_2 \lambda_1 \lambda_3$	{2,3} $\mu_2 \lambda_2 \lambda_3$

Table 3

that the two rows are proportional. Thus, our multiplicative Poisson model has -also for $k > 3$ - the following property:

For any three items, say j_1, j_2 and j_3, consider the subpopulation of subjects who responded to exactly one of the two items j_1 and j_2. Classify these subjects according to two criteria, namely

1) response to j_1 or response to j_2, and

2) response or no response to j_3.

This gives a 2 × 2 - table like table 4, and according to our model the counts in this table are Poisson distributed with

	response to j_1 but not to j_2	response to j_2 but not to j_1
no response to j_3		
response to j_3		

Table 4

parameters expressible as the product of a row parameter and a column parameter. This can be tested by the likelihood ratio test, or, in case of small counts, by Fisher's exact test.

The advantage of this "subtest" of the model is that it has a very simple intuitive interpretation, which makes it well suited for model control when specific items are under suspicion for being incongruous (or "too congruous", cfr. section 4). As an illustration of this, consider table 5, where the three items are taken to be the three questions suggested in the imagined example of section 4. Having classified a large

	knowing the years of world war 2, but unable to tell what a beaver eats	knowing what a beaver eats, but unable to tell when world war 2 took place
unable to tell how a bear spends the winter		
knowing how a bear spends the winter		

Table 5

number of children in this manner (omitting those who answer none or both of the two first questions correctly), we would certainly expect to find a significant interaction in this table. This is so because the classification into the two columns can be regarded as a rough classification of children into a group of zoologically informed children and a group of historically informed children (excluding a - probably large - remainder group), and acquaintance with the habits of bears is almost certainly more common in the first group than in the second.

It is possible to set up more general two-way contingency tables for which the hypothesis of independence is implied by our multiplicative Poisson model. The most general form of such a table can be described as follows: Let A_0 denote a (fixed) subset of the item set $\{1,\dots,k\}$. For some r $(0 < r < \# A_0)$, consider the subpopulation of subjects who responded to exactly r of the items in A_0. Now, cross-classify this subpopulation according to two criteria, namely one criterion based entirely on the response pattern within A_0, and one criterion based on the responses to the remaining items (the last classification may typically be according to raw score, since the raw score is a function of the responses to items not in A_0, when the number of responses to items in A_0 is fixed). It is easy to show that the multiplicative Poisson model of section 3 implies independence in any such two-way table. Table 4 is a special case of this, with $A_0 = \{j_1, j_2\}$. As a more sophisticated example, consider table 6, where we take again $A_0 = \{j_1, j_2\}$, with the first classification as in table 4, but with a second classification

	response to j_1 but not to j_2	response to j_2 but not to j_1
raw score = 1		
2		
⋮		
k - 1		

Table 6

by raw score. The interpretation of a deviation from independence in this thble is similar to what was said in connection with table 4 and 5, except that the role of item j_3 is now taken over by the whole set of remaining items. Thus, our conclusion of a significant interaction in table 6 would be that one of the two items j_1 and j_2 is "more similar" to the main body of items than the other, typically because the latter is "incongruous".

6. Control of the Rasch model.

The statistical tests suggested in section 5 have been explained as tests in the multiplicative Poisson model which, apart from the randomization of the total number n of subjects, is equivalent to an extension of the Rasch model with random row effects. However, as we shall see now, these tests do also make sense as conditional procedures for control of Rasch's original model. Intuitively, this is not surprising, because the interpretation of the row effects as fixed or random is very often a matter of taste. Obviously, there exists no test procedure (independent of the ordering of subjects) that can decide whether the row effects "are" random or fixed. Consequently, any test for goodness of fit of the random effects model (and, in particular, any goodness of fit test for the multiplicative Prisson model) should also be applicable to the fixed effects model.

The likelihood-ratio test of the 2k-dimensional multiplicative Poisson model against the full 2^k-dimensional model can be regarded as a test of the conditional Rasch model in the following sense: Given the raw scores $r_1,\dots,r_n$, consider the 2^k-contingency table (n_A) of section 3. Obviously, the distribution of the n_A's under the conditional Rasch model is equal to the distribution obtained under the multiplicative Poisson model by conditioning on the marginals $n_r = \sum_{A:\#A=r} n_A$, $r = 0, 1, \dots, k$. Now, in a short (but convenient) notation, the likelihood-ratio statistic derived in section 5 can be written (with P for probabilities under the multiplicative Poisson model) as

$$\frac{\max\limits_{\text{2k-dim. model}} P((n_A))}{\max\limits_{\text{full model}} P((n_A))} =$$

$$\frac{\max\limits_{\text{2k-dim. model}} P((n_A)\,|\,n_0\,,n_1,\dots,n_k)P(n_0\,,n_1,\dots,n_k)}{\max\limits_{\text{full model}} P((n_A)\,|\,n_0\,,n_1,\dots,n_k)P(n_0\,,n_1,\dots,n_k)} =$$

$$\frac{\max\limits_{\text{2k-dim. model}} P((n_A) \mid n_0, n_1, \ldots, n_k)}{\max\limits_{\text{full model}} P((n_A) \mid n_0, n_1, \ldots, n_k)}$$

The last identity follows from the fact that the maximum-likelihood estimates of the parameters in the Poisson distributions of $n_0, n_1, \ldots, n_k$ are the same in the 2k-dimensional model and the full model, namely those corresponding to perfect fit. We shall not go into further details with this argument, since it is quite similar to arguments known from the theory of contingency tables, showing, for example, that the likelihood ratio test of independence in a two-way table is the same in the Poisson model and the multinomial model obtained from the Poisson model by conditioning on the row marginals. Our conclusion in the present situation is that the likelihood-ratio statistic for test of the 2k-dimensional multiplicative Poisson model against the 2k-dimensional full model is equal to the likelihood-ratio statistic for test of the conditional Rasch model against the corresponding "full conditional model". By the "full conditional model" we mean, here, the model specified by a $\binom{k}{r}$-dimensional probability vector $(p_A \mid A \subseteq \{1, \ldots, k\}, \# A = r)$ for each $r = 0, 1, \ldots, k$, giving the (arbitrary) distribution of the response configuration for a member of score group r.

Similarly, the likelihood-ratio test of our multiplicative Poisson model against the model specified by a separate set of item parameters for each score group, is equivalent to a likelihood-ratio test in the conditional Rasch model, namely the test against the product of $k-1$ conditional Rasch models (with separate sets of item parameters), one for each non-trivial score group. In section 5, this test was ascribed to Martin-Löf (1970) and Andersen (1973a , 1973b) , but it would, perhaps, be more correct to say that Martin-Löf introduced the test in the model referred to here as the extended random model (or even in the multiplicative Poisson model, since Martin-Löf also imposes a Poisson distribution on n in order to simplify the computations). Andersen derived the test in the Rasch model (and its generalization to several response categories, cfr. section 7) as a special case of a conditional goodness of fit test. Control plots, based on a similar idea of comparison of estimates within score groups, were suggested by Rasch (1960).

The test of independence in a 2 × 2-table (table 4) is also equivalent to a conditional test in the Rasch model. Indeed, let x_1 , x_2 and x_3 denote the responses of an arbitrary subject to items 1 , 2 and 3. According to the Rasch model, we have (by straightforward computations)

$$P(x_1 = 1 \mid x_1 + x_2 = 1 , x_3 = 0) =$$

$$P(x_1 = 1 \mid x_1 + x_2 = 1 , x_3 = 1) = \frac{\lambda_1}{\lambda_1 + \lambda_2} .$$

Hence, if we classify the subpopulation of subjects with $x_1 + x_2 = 1$ and $x_3 = 0$ (i.e. the subpopulation of subjects counted in the upper row of table 4) according to their reaction on j_3, then the (conditional) classification probabilities are independent of the subject parameters, and they coincide with the (conditional) classification probabilities for the same classification applied to the subpopulation given by $x_1 + x_2 = 1$ and $x_3 = 1$ (i.e. the subpopulation of subjects occurring in the lower row of table 4).

It follows, under the conditional distribution in the Rasch model, given the sizes of the two subpopulations, that the two rows of table 4 can be regarded as pairs of counts of success and failure in two binomial distributions with the same probability parameter. It is well known that the likelihood-ratio test for identity of these two parameters coincides with the usual likelihood ratio test of independence in the 2 × 2-table.

Similar interpretations can be given to the tests of independence in the more complicated two-way contingency tables mentioned at the end of section 5.

7. Extension to several response categories.

The present study is restricted to the case of binary responses. It should be mentioned, however, that (apart from overwhelming notational difficulties) all results of this paper can be generalized and/or modified to cover the case of several response categories. To indicate the idea, we shall give a summary of what things are like in the case of three response categories.

As in section 1, we have an $n \times k$ - table of reponses x_{ij}, but now the x_{ij}'s can take three possible values, say 1,2 and 3. Rasch's model for this case is

$$P(x_{ij} = x) = \frac{\lambda_j^{(x)} \xi_i^{(x)}}{\lambda_j^{(1)} \xi_i^{(1)} + \lambda_j^{(2)} \xi_i^{(2)} + \lambda_j^{(3)} \xi_i^{(3)}} .$$

Thus, each subject and each item is described by three parameters (in the binary case we had only one for each, but this was merely to avoid an unnecessary overparametrization. Similarly, we might have taken $\lambda_j^{(1)} = \xi_i^{(1)} = 1$ in the present case, but we prefer to emphasize the symmetry between responses here).

By the <u>raw score</u> for subject i we mean the triple $(r_i^{(1)}, r_i^{(2)}, r_i^{(3)})$ with $r_i^{(1)} + r_i^{(2)} + r_i^{(3)} = k$, where $r_i^{(x)}$ denotes the number of items to which subject i gave the response x. Just as in the binary case, the subject parameters can be eliminated by conditioning on the raw scores, and the conditional maximum-likelihood estimates of the item parameters $\lambda_j^{(x)}$ have desirable asymptotic properties, see Andersen (1973a).

The analogue of the 2^k-contingency table of section 3 is formed by cross-classification of the subjects according to their responses to the k items. This gives a 3^k-contingency table. As our model for this, we may take a multiplicative Poisson model, with one factor for each of the k items and with the raw score as an additional factor. Thus, for $(a_1,\ldots,a_k) \in \{1,2,3\}^k$, the count $n_{(a_1,\ldots,a_k)}$ of the corresponding cell is assumed to be Poisson distributed with parameter

$$E\, n_{(a_1,\ldots,a_k)} = \lambda_1^{(a_1)} \cdots \lambda_k^{(a_k)} \, \mu_{(r^{(1)},r^{(2)},r^{(3)})} ,$$

where $(r^{(1)},r^{(2)},r^{(3)})$ denotes the raw score for response pattern $(a_1,\ldots,a_k)$. This model is related to the conditional Rasch model with three response categories in exactly the same way as in the case of binary responses.

Acknowledgement.

I am grateful to Steffen L. Lauritzen for comments on the paper.

References.

Andersen, E. B. (1973a). Conditional inference and models for measuring, Mentalhygiejnisk Forlag, Copenhagen.

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

Andersen, E. B. (1980). Discrete statistical models with social science applications, North-Holland.

Cox, D. R. (1970). The analysis of binary data, Methuen & Co.

Finney, D. J. (1971). Probit analysis, Cambridge University Press.

Haberman, S. J. (1978). Analysis of qualitative data, vol. I, Academic Press.

Haberman, S. J. (1979). Analysis of qualitative data, vol. II, Academic Press.

Martin-Löf, P. (1970). Statistiska modeller, lecture notes by Rolf Sundberg, Institutet för försäkringsmatematik och matematisk statistik vid Stockholm universitet (in Swedish).

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, J.R.S.S. (A) 135, 370-384.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests, Danmarks Pædagogiske Institut, Copenhagen.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology, Proc. fourth Berkely symp. on math. stat. and prob., vol. IV, 321-333.

PREPRINTS 1979

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Edwards, David: Large Sample Tests for Stationarity and Reversibility in Finite Markov Chains.

No. 2 Andersson, Steen A. : Distribution of Maximal Invariants Using Proper Action and Quotient Measures.

No. 3 Johansen, Søren: A Note on the Welch-James Approximation to the Distribution of the Residual Sum of Squares in a Weighted Linear Regression.

No. 4 Björnsson, Ottó: Four Simple Characterizations of Standard Borel Spaces.

No. 5 Johansen, Søren: Some Comments on Robustness

No. 6 Hald, Anders: T.N. Thiele's Contributions to Statistics.

No. 7 Jacobsen, Martin: Markov Chains: Birth and Death Times with Conditional Independence.

PREPRINTS 1980

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Olkin, Ingram and Væth, Michael: Maximum Likelihood Estimation in a Two-way Analysis of Variance with Correlated Errors in One Classification.

No. 2 Skovgaard, Ib M.: Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions.

No. 3 Asmussen, Søren: Equilibrium Properties of the M/G/1 Queue.

No. 4 Johansen, Søren and Keiding, Susanne: A Family of Models for the Elimination of Substrate in the Liver.

No. 5 Skovgaard, Ib M.: Edgeworth Expansions of the Distributions of Maximum Likelihood Estimators.

No. 6 Tjur, Tue: A Connection Between Rasch's Item Analysis Model and a Multiplicative Poisson Model.