A GENERALISED LINEAR LOGISTIC
TEST MODEL (GLLTM)

Reinhold HATZINGER

Forschungsbericht/
Research Memorandum No. 172

June 1982
Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich des Autors und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden.
## Contents:

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract/Zusammenfassung</td>
<td>2</td>
</tr>
<tr>
<td>1. Estimating the parameters of the Rasch-model and the linear logistic test-model</td>
<td>3</td>
</tr>
<tr>
<td>2. Reparameterising the personparameters in the LLTM - a generalised linear logistic testmodel (GLLTM)</td>
<td>10</td>
</tr>
<tr>
<td>References</td>
<td>15</td>
</tr>
</tbody>
</table>
A GENERALISED LINEAR LOGISTIC TEST MODEL (GLLTM)

Abstract:
An extension of the Rasch-model was introduced by G.Fischer (1972) by imposing a linear structure on the itemparameters. The two corresponding models are briefly presented. Since many applications additionally require such a structure for parameters describing the individuals under observation a model is proposed where both types of parameters can be restricted by taking into account the effect of covariates. Besides a discussion of conditional and unconditional ML-methods an asymptotic procedure is suggested for simplifying parameter estimation.

Zusammenfassung:
1. ESTIMATING THE PARAMETERS OF THE RASCH-MODEL
AND THE LINEAR LOGISTIC TEST-MODEL

In experiments frequently the following situation arise: a number of objects is exposed to certain conditions and the interest is whether a reaction specified can be observed or not. The outcome is then simply coded by assigning 0 or 1 to the object according to the response. This corresponds to the assumption of observing the realisations of a sequence of binary random variables. For illustrative purpose the objects might be persons and the conditions are items of an ability-test where the correct solution of an item is coded by 1. Furthermore a quantification of the difficulty of the items and the ability of the subjects is required. (As a convention the terms "item-" and "individual parameters" shall be used for the corresponding quantities in the following.)

Denoting the effect of item j by $\lambda_j$ (j=1,...,k) and the ability of subject i by $\xi_i$ (i=1,...,n) the probability $p_{ij}$ for subject i to have a correct solution on item j is according to Rasch's well-known proposal given as

$$p_{ij} = \frac{\exp(\xi_i + \lambda_j)}{[1 + \exp(\xi_i + \lambda_j)]}.$$

As one of his main results Rasch formulated a conditional maximum likelihood (CML) for (1) where by conditioning on the row or column marginals of the datamatrix $Y = (y_{ij})$ either the $\xi_i$'s or the $\lambda_j$'s cancel out. This can be shown by the following arguments: Let $A_i = A$ be the realisation of a sequence of Bernoulli-variables $Y_{ij} = y_{ij}$ (j=1,...,k) where $A_i$ is the response pattern of subject i. The probability $P(A_i = A)$ for observing exactly
this sequence is

\[ P(A_i = A) = \exp(\xi_i t_i + \sum_j \lambda_j y_{ij}) \prod_j [1 + \exp(\xi_i + \lambda_j)]^{-1} \]

\( t_i = \sum_j y_{ij} \) is the marginal total for subject \( i \). The probability for observing \( t_i \)

\[ P(T_i = t_i) = \sum_{B|i} P(A_i = B) \]

is the sum over all possible response patterns \( B \) that contain exactly \( t_i \) elements (i.e. correct solutions or 1's) for the probability of sequence \( A_i \). Now assuming \( t_i \) as given the conditional likelihood for observing \( A_i \) given the marginal \( t_i \) is

\[ P(A_i = A | T_i = T) = \frac{P(A_i = A)}{\sum_{B|i} P(A_i = B)} = \frac{\exp(\xi_i t_i) \exp(\sum_j \lambda_j) \prod_j [1 + \exp(\xi_i + \lambda_j)]^{-1}}{\exp(\xi_i t_i) \sum_{B|i} \exp(\sum_j \lambda_j) \prod_j [1 + \exp(\xi_i + \lambda_j)]^{-1}} = \frac{\exp(\sum_j \lambda_j) / \sum_{B|i} \exp(\sum_j \lambda_j)}{\sum_{B|i} \exp(\sum_j \lambda_j)} = \frac{\exp(\sum_{j \in A} \lambda_j)}{\lambda_i} \cdot f_{t_i}^*(\lambda_1, \ldots, \lambda_k) \]

where \( f_{t_i}^* \) is the elementary symmetric function of order \( t_i \).
By multiplying (3) over all $i$ the global likelihood $L_c$ (c for conditional) results

\begin{equation}
L_c(\lambda_1, \ldots, \lambda_k) = \prod_i \exp(\sum_j \lambda_j) \cdot \left[ \gamma_{t_i}^3(\lambda_1, \ldots, \lambda_k) \right]^{-1} = \\
= K \frac{\exp(\sum \lambda_j s_j)}{\prod_{r=1}^k \gamma_r(\lambda_1, \ldots, \lambda_k)^n_r}
\end{equation}

$K$ denotes a combinatorial factor reflecting the total number of data matrices that satisfy the row marginals $t_i = r$ ($r=1, \ldots, k-1$) and the column totals $s_j = \sum_i y_{ij}$, for all $i$ and $j$.

Overparameterisation of (4) by one parameter can easily be avoided e.g. by restricting the $\lambda_j$'s to $\sum \lambda_j = 0$.

Maximising (4) yields the likelihood equations for the $\lambda_j$'s

\begin{equation}
\frac{\partial \log L_c}{\partial \lambda_\delta} = -s_r + \sum_r n_r \exp(\lambda_\delta) \cdot \gamma_{r-1}^{(3)}(\lambda) \cdot \gamma_r(\lambda)^{-1} = 0,
\end{equation}

$\delta = 1, \ldots, k-1$,

$\gamma_{r-1}(\lambda) = \gamma_{r-1}^{(x)}(\lambda_1, \ldots, \lambda_{r-1}, \lambda_{r+1}, \ldots, \lambda_k)$.

Advantages for using (4) instead of the full (unconditional) likelihood for parameter estimation have extensively been discussed (cf. ANDERSEN 1973a, 1980; FISCHER 1974) and shall not be treated here in further detail.
One of the several extensions of the Rasch-model (RM) was introduced by G.H. FISCHER in 1972 (see also FISCHER, 1974) where he formulated the linear structure

\[ \lambda_j = \sum_{q=1}^{m} a_{jq} \alpha_q + c \]  

for the item parameters. Inserting (6) into (1) provides the model

\[ P_{ij} = \frac{\exp(\delta_i + \sum_{q} a_{jq} \alpha_q + c)}{1 + \exp(\delta_i + \sum_{q} a_{jq} \alpha_q + c)} \]

which was according to COX's terminology called linear logistic test-model (LLTM). The basic idea leading to (7) was to linearise the common structure of the items (if such a structure can be assumed) or equivalently to reduce the dimensionality of the space spanned up by the \( \lambda \)'s.

The corresponding CML is

\[ L_c(\alpha_1, \ldots, \alpha_m) = K \frac{\exp(\sum_{j} s_j \sum_{q} a_{jq} \alpha_q)}{\prod_{r} \gamma_r(g(\alpha))^{n_r}} = K \frac{\exp(\sum_{j} s'_q \alpha_q)}{\prod_{r} \gamma_r(g(\alpha))^{n_r}} \]

\[ s'_q = \sum_j s_j a_{jq} \]

Taking partial derivatives with respect to \( \alpha_q \) of the log \( L_c \)

\[ \log L_c(\alpha_1, \ldots, \alpha_m) = \ln K + \sum_{q} s'_{q} \alpha_{q} - \sum_{r} n_r \gamma_{r}(g(\alpha)) \]
yields the estimation equations

\[(10) \sum_j a_j \left[ s_j - \sum_r n_r \left\{ \lambda_j^{(r)} \exp(\lambda_j) / \lambda_j^{(r)} \right\} \right] = 0.\]

It should be noted that the \(\lambda_j\) and the \(\lambda_j^{(r)}\) in (10) are functions of the \(\lambda_j\)'s which themselves are functions of the \(\alpha_q\)'s.

As usual a LR test can be examined for evaluating the adequacy of the restriction (6) by comparing the restricted model (7) with the full model (1). The corresponding statistic is

\[(11) -2 \ln \left[ \frac{L_c(\alpha_1, \ldots, \alpha_m)}{L_c(\lambda_1, \ldots, \lambda_k)} \right].\]

approximately \(\chi^2\)-distributed with \(df = k - m - 1\).

However (9) does not check the general structure of (1) or (6) respectively. This might be done by either using graphical methods or any of several goodness-of-fit tests for the RM. (For detailed discussion see FISCHER, 1974 or HATZINGER, forthcoming)

Since (4) and (8) obviously don't contain information for estimating the person parameters \(\xi_i\)'s the question arises how to proceed if additional quantification of individual effects on the experiment mentioned above is required (i.e. the experiment is an ability test and there is necessity for comparing the subjects under observation).
Models (1) and (7) are symmetrical with respect to both types of parameters, thus usage of conditional methods according to (4) and (8) can be considered as providing reasonable results. The corresponding likelihoods coincide to

\[(12) \quad L_c(\xi_1, \ldots, \xi_n) = K \cdot \exp(\sum_i \xi_i t_i) \cdot \prod_j \left[ \frac{1}{n_j} \exp(\xi_j) \right]^{-1}\]

which by maximisation leads to estimation equations analogous to (5).

Unfortunately practical applications of (12) fail when large \( n \) have been sampled (a condition for sufficient accuracy of parameter estimates for the \( \lambda_j \)'s) since numerical solutions require heavy time- and space-consuming procedures. Thus an alternative approach is provided by usage of unconditional arguments (cf. FISCHER, 1974). The unconditional likelihood \( L_u \) for (1) is given as

\[(13) \quad L_u(\xi_i, \lambda_j \text{ for all } i,j) = \frac{\exp(\sum_i \xi_i t_i + \sum_j \lambda_j s_j)}{\prod_i \prod_j \left[ 1 + \exp(\xi_i + \lambda_j) \right]}\]

and for (7) accordingly

\[(14) \quad L_u(\xi_i, \alpha_q \text{ for all } i,q) = \frac{\exp(\sum_i \xi_i t_i + \sum_q \alpha_q s_q^i)}{\prod_i \prod_q \left[ 1 + \exp(\xi_i + \sum_q \alpha_q) \right]}\]

Differentiation of the logarithm of (13) with respect to \( \xi_i \)
yields the estimation equations for the person parameters

\[(15) \quad t_i = \sum_j \exp(\xi_i + \lambda_j) / \left[ 1 + \exp(\xi_i + \lambda_j) \right]. \]

Since the \( t_i \)'s are the sufficient statistics for the \( \hat{\xi}_i \)'s the number of different \( \hat{\xi}_i \)'s is restricted to \( k-1 \). (In cases where \( t_i = 0 \) or \( k \) respectively no information is left for estimating the corresponding \( \hat{\xi}_i \)'s.)

If \( n \) is large compared to \( k \) the estimates \( \hat{\lambda}_j \) having used CML-methods should be so close to the true values of \( \lambda_j \) that they can be treated as known constants ignoring their standard errors. It seems reasonable therefor to insert the \( \hat{\lambda}_j \)'s for the \( \lambda_j \)'s in (15), which leads to explicit solutions for the \( \hat{\xi}_i \)'s by simply iterating (15) until

\[ \hat{\xi}_t - \hat{\xi}_{t-1} < e \] for all \( t \) (\( t=1,..k-1 \)), where \( e \) is chosen sufficiently small. The same arguments apply for the LITM.
2. REPARAMETERISING THE PERSON PARAMETERS IN THE LLTM - A GENERALISED LINEAR LOGISTIC TESTMODEL (GLLTM)

In (7) the LLTM resulted by extension of (6), namely by imposing the linear reparameterisation (6) on the item-parameters of the RM. A generalised model can be formulated by inserting

\[
\xi_i = \sum_p b_{ip} \beta_p + c' \quad p = 1, \ldots, o
\]

into (7), which leads to

\[
p_{ij} = \frac{\exp(\sum_p b_{ip} \alpha_p + \sum_q a_{iq} \alpha_q + c'')}{1 + \exp(\sum_p b_{ip} \alpha_p + \sum_q a_{iq} \alpha_q + c'')}
\]

\begin{align*}
i & = 1, \ldots, n \\
j & = 1, \ldots, k \\
p & = 1, \ldots, o \\
q & = 1, \ldots, m
\end{align*}

By using conditional arguments for obtaining ML-equations the same methods obviously apply as presented in the previous section. Thus the \(\alpha_q\)'s and the \(\beta_p\)'s might be estimated by the corresponding CMLs. Again in cases of large \(n\) the order of the elementary symmetric functions for the \(\xi_i\)'s and the \(\beta_p\)'s respectively is very high and hence computations are probable to fail for reasons of computer capacity.
A numerical solution for the \( \hat{\beta}'s \) can be obtained in analogy to (15) by maximising the likelihood for (17) and inserting the \( \hat{\lambda}'s \) or \( \hat{\alpha}'s \) already estimated by means of CML. The system of estimation equations

\[
\frac{\partial \ln L_u(\beta, \lambda)}{\partial \beta_p} = t'_p - \sum_i \sum_j b_{ip} \exp(\sum_p b_{ip} \beta_p + \lambda_j) \over 1 + \exp(\sum_p b_{ip} \beta_p + \lambda_j)
\]

with \( t'_p = \sum_i t_i b_{ip} \) can then iteratively be solved e.q. by the Newton-Raphsen method.

By using the empirical logistic transform an alternative approximate solution might be given by using the following arguments introduced by J.BERKSON as the method of "minimum logit \( \chi^2 \)"-estimation of the bio-assay and which in a somewhat different context have been presented by D.R.CO(1970):

Suppose a grouping of binary observations into sets \( i \) where all trials \( j \) \( (j=1, \ldots, k) \) within set \( i \) having the same probability \( p_i=p_{i1}=\ldots=p_{ik} \). Given \( t_i \) successes in \( k \) trials then

\[
p_i = E(t_i/k) \quad \text{and} \quad \text{var}(t_i/k) = p_i(1-p_i)/k
\]

Now \( \Psi(p) \), a column vector with elements \( \Psi(p_i) \) and denoting the logit transform of the probabilities \( p_i \), \( (p)=\ln(p/(1-p)) \), specifies the linear model

\[
\Psi(p) = B \quad \text{or} \quad \Psi(p_i) = \sum_p b_{ip} \beta_p
\]

For large \( k \) \( \Psi(p_i) \) is nearly normally distributed with expectation

\[
z_i = E(\Psi(t_i/k))
\]
and variance

\[(21) \quad V_i = \left[ \psi'(p_i) \right]^2 p_i (1 - p_i) / k \]

if none of the \(p_i\)'s is very close to 0 or 1. (In practical applications care must only be taken if very extreme regions are of special interest.) As \(k \to \infty\) asymptotic mean and variance can consistently be estimated by

\[(22) \quad \hat{z}_i = \ln \left( \frac{t_i}{(k-t_i)} \right), \quad \text{and} \]

\[(23) \quad \hat{V}_i = \left\{ \ln \left[ \frac{t_i/k}{1 - t_i/k} \right] \right\}^2 \frac{t_i}{k} \left( 1 - t_i/k \right) / k = \]

\[= k / (t_i (k-t_i)) \]

This leads to the possibility for either using OLS-methods if the \(V_i\)'s do not differ very much or weighted least-squares analyses with empirical weights \(\sqrt{V_i}\) to estimate the \(\beta\)'s. Thus

\[(24) \quad E \left( \frac{z_i}{\sqrt{V_i}} \right) = \sum_p \left( \frac{b_{ip}}{\sqrt{V_i}} \right) \beta_p \]

can be treated as having unit variance when ignoring random errors in \(b_{ip}/\sqrt{V_i}\) being a second order effect.
It should be noted that the results hold as well for other transforms such as probit, angular or complementary log-log. The properties of the estimates \( \hat{\beta} \) were discussed by W.F.TAYLOR (1953) who provided the proof that the estimates are regular best asymptotically normal (R.B.A.N.) in the sense of Neyman and thus have asymptotically the same properties as ML-estimates. Furthermore it was found by extensive trials that the method provides good estimates even if \( k \) is quite small (J.BERKSON 1953, 1955).

Applying these considerations to the GLLTM (17) equation (24) becomes

\[
(25) \quad E(z_i/\sqrt{V_i}) = E \left[ \ln(t_i/(k-t_i)) \cdot \{k/(t_i(k-t_i))\}^{\frac{1}{2}} \right] = \sum b_{ip} \sqrt{t_i/(k-t_i)} \beta_i \]

Denoting the \((ix1)\)-vector of \( E(z_i/\sqrt{V_i}) \) by \( \chi \), the \((ixp)\)-design matrix of \( b_{ip} \cdot \sqrt{V_i} \) by \( B \) and the \((px1)\)-parameter vector by \( \beta \) the regression is

\[
(26) \quad \chi = B\beta + \epsilon, \quad \epsilon \sim N(0, I). \]

The WLS estimation is then as usually given by

\[
(27) \quad \hat{\beta} = (B'B)^{-1}B'\chi. \]

Well-known F-statistics with corresponding degrees of freedom can be used for evaluating the linear restriction (16) (16), or to test the effect of a single parameter.
Unfortunately the requirement for constant probabilities
\( p_i = p_{i1} = \ldots = p_{ik} \) over the whole sequence is not met
apriori. Estimation of the \( \beta \)'s using the empirical
logistic transform is consistent only if \( \lambda_1 = \ldots = \lambda_k \),
which in most cases seems to be an unrealistic assumption.
Thus the \( \beta \)'s estimated by (27) might be biased by a
variation of the \( \lambda \)'s. However, the adjustment

\[
(28) \quad \psi(\hat{p}_i) = u_1 + u_2 \ln(t_i/(k-t_i)) = u_1 + u_2 \psi(\hat{p}_i)
\]

on the logistic scale should suffice. The resulting
probability \( \hat{p}_i \)

\[
(29) \quad \hat{p}_i = \frac{e^{u_1 \hat{p}_i + u_2}}{e^{u_1 \hat{p}_i + u_2} + (1-\hat{p}_i)^u_2}
\]

has then the interpretation as the probability for subject \( i \)
to have correct solutions independent of the difficulty of
special items. Individual difficulties of items increase
or decrease \( \hat{p}_i \).

\( u_1 \) and \( u_2 \) are calculated by (i) using (15) to obtain \( \hat{t}_i \),
(ii) finding \( \hat{p}_i \) by \( \hat{p}_i = \exp(\hat{t}_i)/(1+\exp(\hat{t}_i)) \) and (iii) in-
serting \( \hat{p}_i \) into (28). Estimated variances of the transformed
\( \hat{p}_i \)'s required for the WLS-analysis (26) are given by
\( \hat{V}_i = u_2^2 \hat{V}_i \).
REFERENCES


FISCHER, G.H.: Einführung in die Theorie psychologischer Tests. 1974, Huber, Bern


