A GENERALISED LINEAR LOGISTIC TEST MODEL (GLLTM)

Reinhold HATZINGER

Forschungsbericht/
Research Memorandum No. 172

June 1982

Contents:

page

Abstract/Zusammenfassung 2

1. Estimating the parameters of the Rasch-model and the linear logistic test-model 3

2. Reparameterising the personparameters in the LLTM - a generalised linear logistic testmodel (GLLTM) 10

References 15

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:
G.Fischer(1972) formulierte eine Erweiterung des Rasch Models durch Einführen linearer Nebenbedingungen über die Itemparameter. Die beiden entsprechenden Modelle werden kurz dargestellt. Da in vielen Anwendungsfällen eine derartige Struktur auch für die Parameter erforderlich ist, die die Beobachtungseinheiten (Versuchspersonen) beschreiben, wird ein Modell vorgeschlagen, in dem beide Typen von Parametern linear reparametrisiert werden können. Neben einer Diskussion bedingter und unbedingter ML-Methoden wird ein asymptotisches Verfahren dargestellt, das die Parameterschätzung vereinfacht.

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 λ_j (j=1,..,k) and the ability of subject i by ξ_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

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

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 ξ_i's or the λ_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

$$(2) \qquad P(A_i=A) = \exp(\xi_i t_i + \sum_j \lambda_j y_{ij}) \prod_j \left[1 + \exp(_i + _j)\right]^{-1}$$

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

$$P(T_i=t_i) = \sum_{B|t_i} P(A_i=B)$$

is the sum over all possible response patterns B that contain exactly t_i elements (i.e. corrects 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

$$(3) \qquad P(A_i=A|T_i=T) = \frac{P(A_i = A)}{\sum_{B|t_i} P(A_i = B)} =$$

$$= \frac{\exp(\xi_i t_i)\exp(\sum_{j\in A}\lambda_j) \cdot \prod_j [1 + \exp(\xi_i + \lambda_j)]^{-1}}{\exp(\xi_i t_i)\sum_{B|t_i}\exp(\sum_{j\in B}\lambda_j) \cdot \prod_j [1 + \exp(\xi_i + \lambda_j)]^{-1}} =$$

$$= \exp(\sum_{j\in A}\lambda_j)/\sum_{B|t_i}\exp(\sum_{j\in B}\lambda_j) =$$

$$= \exp(\sum_{j\in A}\lambda_j)/\gamma_{t_i}(\lambda_1,\ldots,\lambda_k)$$

where γ_{t_i} is the elementary symetric function of order t_i.

By multiplying (3) over all i the global likelihood L_c (c for conditional) results

$$(4) \qquad L_c(\lambda_1,\ldots,\lambda_k) = \prod_i \exp(\sum_{j\in A_i}\lambda_j)\cdot[\gamma_{t_i}(\lambda_1,\ldots,\lambda_k)]^{-1} =$$

$$= K \frac{\exp(\sum_j \lambda_j s_j)}{\prod_{r=1}^{k-1} \gamma_r(\lambda_1,\ldots,\lambda_k)^{n_r}}.$$

K denotes a kombinatorial 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.q. by restricting the λ_j's to $\sum_j \lambda_j = 0$.

Maximising (4) yields the likelihood equations for the λ_j's

$$(5) \qquad \frac{\partial \log L_c}{\partial \lambda_\delta} = -s_\delta + \sum_r n_r \exp(\lambda_\delta)\cdot \gamma_{r-1}^{(\delta)}(\underset{\sim}{\lambda})\cdot \gamma_r(\underset{\sim}{\lambda})^{-1} = 0,$$

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

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

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

$$(6) \qquad \lambda_j = \sum_{q=1}^{m} a_{jq}\alpha_q + c$$

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

$$(7) \qquad p_{ij} = \frac{\exp(\xi_i + \sum_q a_{jq}\alpha_q + c)}{1 + \exp(\xi_i + \sum_q a_{jq}\alpha_q + c)} \, .$$

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 λ's.

The corresponding CML is

$$(8) \qquad L_c(\alpha_1,\ldots,\alpha_m) = K \frac{\exp(\sum_j s_j \sum_q a_{jq}\alpha_q)}{\prod_r \gamma_r(g[\underset{\sim}{\lambda}])^{n_r}} = K \frac{\exp(\sum_q s'_q \alpha_q)}{\prod_r \gamma_r(g[\underset{\sim}{\lambda}])^{n_r}}$$

$$s'_q = \sum_j s_j a_{jq}$$

Taking partial derivatives with respect to α_γ of the log L_c

$$(9) \qquad \log L_c(\alpha_1,\ldots,\alpha_m) = \ln K + \sum_q s'_q\alpha_q - \sum_r n_r \gamma_r(g(\underset{\sim}{\lambda}))$$

yields the estimation equations

$$(10) \qquad \sum_j a_{j\delta} \left[s_j - \sum_r n_r \left\{ \gamma^{(\delta)}_{r-1}(g(\underset{\sim}{\lambda})) \cdot \exp(\lambda_\delta) / \gamma_r(g(\underset{\sim}{\lambda})) \right\} \right] = 0 \; .$$

It should be noted that the γ_r and the $\gamma^{(\delta)}_{r-1}$ in (10) are functions of the λ_j's which themselves are functions of the α_q's.

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

$$(11) \qquad -2 \ln \left[L_c(\alpha_1,..,\alpha_m)/L_c(\lambda_1,..,\lambda_k) \right] ,$$

approximately χ^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 detailled discussion see FISCHER,1974 or HATZINGER,forthcoming)

Since (4) and (8) obviously don't contain information for estimating the person parameters ξ_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 symetrical 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

$$(12)\quad L_c(\xi_1,\ldots,\xi_n) = K \cdot \exp\left(\sum_i \xi_i t_i\right) \cdot \prod_{s_j}\left[\gamma_{s_j}(\underset{\sim}{\xi})^{n_{s_j}}\right]^{-1}$$

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 λ_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

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

and for (7) accordingly

$$(14)\quad L_u(\xi_i,\alpha_q; \forall\, i,q) = \frac{\exp\left(\sum_i \xi_i t_i + \sum_j \alpha_q s'_q\right)}{\prod_i \prod_j [1 + \exp(\xi_i + \sum_q a_{jq}\alpha_q)}$$

Differentiation of the logarithm of (13) with respect to ξ_i

yields the estimation equations for the person parameters

$$(15) \qquad t_i = \sum_j \exp(\xi_i + \lambda_j)/[1 + \exp(\xi_i + \lambda_j)].$$

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 ξ_i's.)
If n is large compaired to k the estimates $\hat{\lambda}_j$ having used CML-methods should be so close to the true values of λ_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 λ_j's in (15), which leads to explicit solutions for the $\hat{\xi}_i$'s by simply iterating (15) until $\hat{\xi}_t - \hat{\xi}_t^* < e$ for all t (t=1,..k-1), where e is chosen sufficiently small. The same arguments apply for the LLTM.

2. REPARAMETERISING THE PERSONPARAMETERS IN THE LLTM - A GENERALISED LINEAR LOGISTIC TESTMODEL (GLLTM)

In (7) the LLTM resulted by extension of (1), namely by imposing the linear reparameterisation (6) on the item-parameters of the RM. A generalised model can be formulated by inserting

$$(16) \qquad \xi_i = \sum_p b_{ip}\beta_p + c' \qquad\qquad p = 1,\ldots,o$$

into (7), which leads to

$$(17) \qquad p_{ij} = \frac{\exp(\sum_p b_{ip}\beta_p + \sum_q a_{jq}\alpha_q + c'')}{1 + \exp(\sum_p b_{ip}\beta_p + \sum_q a_{jq}\alpha_q + c'')}$$

$i = 1,\ldots,n$
$j = 1,\ldots,k$
$p = 1,\ldots,o$
$q = 1,\ldots,m$

By using conditional arguments for obtaining ML-equations the same methods obviously apply as presented in the previous section. Thus the α_q's and the β_p's might be estimated by the corresponding CMLs. Again in cases of large n the order of the elementary symetric functions for the ξ_i's and the β_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 $\hat{\lambda}_j$'s or $\hat{\alpha}_q$'s already estimated by means of CML. The system of estimation equations

$$(18) \qquad \frac{\partial \ln L_u(\underset{\sim}{\beta},\underset{\sim}{\lambda})}{\partial \beta_p} = t'_p - \sum_i \sum_j \frac{b_{ip} \exp(\sum_p b_{ip}\beta_p + \lambda_j)}{1 + \exp(\sum_p b_{ip}\beta_p + \lambda_j)}$$

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 χ^2"-estimation of the bio-assay and which in a somewhat different context have been presented by D.R.COX(1970): Suppose a grouping of binary observations into sets i where all trials j (j=1,..,k) within set i having the same probability $p_i=p_{i1}=\ldots=p_{ik}$. Given t_i successes in k trials then

$$(19) \qquad p_i = E(t_i/k) \qquad \text{and} \qquad var(t_i/k) = p_i(1-p_i)/k$$

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

$$(20) \qquad \Psi(p) = B \qquad \text{or} \qquad \Psi(p_i) = \sum_p b_{ip}\beta_p \quad .$$

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

$$(21) \qquad z_i = E(\Psi(t_i/k))$$

and variance

$$(21) \qquad V_i = \left[\Psi'(p_i)\right]^2 p_i\ (1 - p_i)\ /\ k$$

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

$$(22) \qquad \hat{z}_i = \ln(t_i/(k-t_i)) , \qquad \text{and}$$

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

$$= k\ /\ (t_i(k-t_i))$$

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 β's. Thus

$$(24) \qquad E(\ z_i/\sqrt{V_i}\ ) = \sum_p (b_{ip}/\sqrt{V_i}\ )\ \beta_p$$

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

$$(25)\qquad E(z_i/\sqrt{V_i}) = E\left[\ln(t_i/(k-t_i)) \cdot \{k/(t_i(k-t_i))\}^{-\frac{1}{2}}\right] = $$

$$= \sum_p \{k/(t_i(k-t_i))\}^{-\frac{1}{2}} b_{ip}\beta_p$$

Denoting the (ix1)-vector of $E(z_i/\sqrt{V_i})$ by $\underset{\sim}{y}$, the (ixp)-design matrix of $b_{ip} \cdot V_i^{-\frac{1}{2}}$ by B and the (px1)-parameter vector by $\underset{\sim}{\beta}$ the regression is

$$(26)\qquad \underset{\sim}{y} = B\underset{\sim}{\beta} + e \ , \quad e \sim N(0,I) \ .$$

The WLS estimation is then as usually given by

$$(27)\qquad \hat{\underset{\sim}{\beta}} = (B'B)^{-1}B'\underset{\sim}{y} \ .$$

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 $p_i = p_{i1} = \dots = p_{ik}$ over the whole sequence is not met apriori. Estimation of the β's using the empirical logistic transform is consistent only if $\lambda_1 = \dots = \lambda_k$, which in most cases seems to be an unrealistic assumption. Thus the β's estimated by (27) might be biased by a variation of the λ's. However, the adjustment

$$(28) \qquad \Psi(\hat{p}_i^*) = u_1 + u_2 \ln(t_i/(k-t_i)) = u_1 + u_2 \Psi(\hat{p}_i)$$

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

$$(29) \qquad \hat{p}_i^* = \frac{e^{u_1} \hat{p}_i^{u_2}}{e^{u_1} p_i^{u_2} + (1-\hat{p}_i)^{u_2}}$$

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 p_i^*.
u_1 and u_2 are calculated by (i) using (15) to obtain $\hat{\xi}_i$, (ii) finding $\hat{p}_i^*$ by $\hat{p}_i^* = \exp(\hat{\xi}_i)/(1+\exp(\hat{\xi}_i)$ and (iii) inserting $\hat{p}_i^*$ 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

ANDERSEN,E.B.: Conditional inference for multiple-choice questionaires. Br.J.Math.Stat.Psych.,1973(a), p.31-44

ANDERSEN,E.B.: A goodness-of-fit test for the Rasch model. 1973(b),Psychometrika 38,Vol.1.

ANDERSEN,E.B.: Discrete statistical models with social science application.1980,N.Holl.P.C.,Amsterdam

BERKSON,J.: A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function.1953, J.Am.Stat.Ass.,48,p.565-599

BERKSON,J.: Maximum likelihood and minimum chi-square estimates of the logistic function.1955,J.Am. Stat.Ass.,50,p.130-162

COX,D.R.: The analysis of binary data.1970,Methuen,London

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

FISCHER,G.H.: Conditional maximum likelihood estimation of item parameters for a linear logistic test model.1972,Res.Bull.9,Inst.f.Psych.,Univ.Vienn.

HATZINGER,R.: Goodness-of-fit tests for the Rasch model - a survey.1982,forthcoming

RASCH,G.: Probabilistic models for some intelligence and attainment tests.1960,Inst.f.Educ.Res.,Copenhagen

TAYLOR,W.F.: Distance functions and regular best asymptotically normal estimates.1953,Ann.Math. Stat.,24,p.85-92