Pearson and Log-likelihood Chi-square Test of Fit for Latent Class Analysis Estimated with Complex Samples

Tihomir Asparouhov and Bengt Muthén

December 22, 2008

1 Introduction

In this note we discuss model fit evaluation for the Latent Class Analysis (LCA) model under complex sampling. Suppose that for the $i-$ individual in the sample we observe r categorical/discrete variables $U_{i1}, ..., U_{ir}$. Suppose that for individual i there exist one unobserved categorical variable C_i, called the latent class variable. The LCA model is described by the following equations

$$P(U_{ij} = l | C_i = k) = p_{lk}$$

$$P(C_i = k) = q_k$$

where p_{lk} and q_k are the model parameters that are estimate under the constraints

$$\sum_l p_{lk} = 1$$

$$\sum_k q_k = 1.$$

For complex samples the LCA model is estimated in Mplus by the Pseudo Maximum Likelihood (PML) algorithm, see Skinner (1989b) and Asparouhov (2005). In Asparouhov and Muthen (2005) it is shown that the likelihood ratio test (LRT) can be conducted with complex sampling using the following chi-square adjustment method. Let H0 and H1 be two nested models with design effect matrices Δ_0 and Δ_1, see section 2.11 in Skinner (1989a), and log-likelihood values L_0 and L_1 respectively. Under the more restrictive H0 model the adjusted test statistic

$$X = \frac{2(L_1 - L_0)}{c} \tag{1}$$

has approximately a chi-square distribution with p_1-p_0 degrees of freedom where p_1 and p_0 are the number of parameters and c is the correction factor computed as follows

$$c = \frac{Tr(\Delta_1) - Tr(\Delta_0)}{p_1 - p_0}. \tag{2}$$

In LCA two typical tests of fit are the Pearson chi-square test

$$X_1 = \sum_{j=1}^{J} \frac{(o_j - e_j)^2}{e_j} \tag{3}$$

and the Log-likelihood test of fit

$$X_2 = \sum_{j=1}^{J} o_j \log(o_j/e_j) \tag{4}$$

where the o_j is the observed quantity in the contingency table of the observed data and e_j is the corresponding LCA estimated quantity. Here J denotes the number of cells in the contingency table for the LCA model. These tests are designed to compare the estimated LCA model against the unrestricted contingency table model. In fact the Log-likelihood test of fit (4) can simply be viewed as the usual LRT test between the LCA and the unrestricted contingency table model. Therefore one can use the Asparouhov and Muthen (2005) method (1) for obtaining a proper log-likelihood test statistic under complex sampling.

Note also that under complex sampling o_j in (34) is not simply the observed quantity but it is computed by

$$o_j = \sum_{U_i \in j-th\ cell} w_i,$$

where w_i is the sampling weight for observation i, because these are the PML estimates of the cell probabilities under the H1 model.

Rao and Scott (1984) show that under complex sampling both X_1 and X_2 have the same asymptotic distribution, which is essentially sums of weighted chi-square distributions with one degrees of freedom, i.e.,

$$\sum \delta_j \chi^2(1) \tag{5}$$

where the weights δ_j are the eigenvalues of the design effect matrix Δ of the additional p_1-p_0 parameters present in the H1 model. This implies that the correction factor (2) used for X_2 can actually be used for X_1 as well. Rao and Thomas (1989) also describe a second order approximation to the distribution (5) which matches not only the mean of the distribution, $Tr(\Delta)$, but also its variance, $2Tr(\Delta^2)$.

The Mplus program computes the trace of the design effect matrix for each model estimated under complex sampling and therefore the LRT testing (1-2) is fairly easy to perform with the Mplus program. However there are certain difficulties specific to the contingency table test of fit. The H1 model can potentially have a much larger number of parameters and the computation of the design effect matrix could be potentially quite intensive. Consider for example a 2-class LCA model with 10 binary variables. The H1 model has 1023 parameters while the H0 model has 21 parameters. The large number of parameters can also lead to matrix inversion problems. Therefore it is important to find a computationally efficient solution. Another complication arises from the fact that the H0 and the H1 model can not be easily expressed in the form where the H1 model is simply expanded model with $p_1 - p_0$ additional parameters, this would be needed to compute the design effect matrix and its eigenvalues for deriving a second order approximation to the chi-square distribution. Another problem that arises here is the fact that the H1 model simply matches the observed values and therefore the corresponding estimated probability parameter will take a value of zero when certain pattern does not occur in the sample. This however is a boundary parameter value which violates the assumptions of the above asymptotic formulas which can lead to poor results in the above approximations. In addition boundary parameters can occur in the H0 model estimation. Finally we have to also acknowledge the fact that the Pearson and the Log-Likelihood test of fit for contingency tables is quite difficult to use when the number of cells is large, for example larger than 5000, even when the sample is a simple random sample (SRS). Therefore we can not expect to obtain good results for large tables even if we provide the correct complex sampling adjustment.

2 First Order Correction

Let us first consider the simple case when there is no cluster, multistage, without replacement or stratified sampling, i.e., only unequal probability of selection with replacement is present in the sample and sampling weights inversely proportional to the probability of selection are provided. The pseudo log-likelihood is given by

$$L = \sum_i w_i L(U_i)$$

where w_i are the sampling weights and $L(U_i)$ is the log-likelihood for the $i-$th individual. The design effect matrix for the model is

$$\Delta = Var(L')(L''^{-1}) \tag{6}$$

where L' and L'' are the first and second derivatives of the log-likelihood L. The variance covariance $Var(L')$ is approximated by

$$Var(L') = \sum_i w_i^2 L'(U_i) L'(U_i)^T. \tag{7}$$

Let's call this the direct method for computing the design effect matrix. Since the number of parameters is relatively small in the H0 model we use the direct method for computing the design effect matrix for the H0 model

We now focus on the computation of the design effect matrix for the H1 model. We are going to compare the direct method described above with two other methods. The first method is given in formula (4.15) in Rao-Thomas (1989)

$$Tr(\Delta_1) = n \sum_{j=1}^{J} \frac{Var(\hat{\mu}_j)}{\hat{\mu}_j} \tag{8}$$

where μ_j is the probability that an observation is in the $j-$th cell of the contingency table and n is the sample size. Under the unequal probability complex sampling

$$Var(\hat{\mu}_i) = \frac{(1-\mu_i)^2 \sum_{U_i \in j-th\ cell} w_i^2 + \mu_i^2 \sum_{U_i \notin j-th\ cell} w_i^2}{(\sum_{i=1}^n w_i)^2}$$

Now we are going to describe an alternative approximate method for computing the H1 design effect matrix. Suppose that we artificially augment the H1 model by an additional cell, which never occurs in the data. Let's call

Table 1: Comparing method for computing the trace of the H1 design matrix

Method	$\mathrm{Tr}(\Delta_1)$
Direct	5.057
Rao-Thomas	5.057
Approximate	5.105

this the $J+1$ cell. Under this augmented model the pseudo H1 log-likelihood simplifies to

$$\sum_{i=1,U_i\in j-th\ cell}^{n} w_i log(\mu_j),$$

i.e., this augmentation is equivalent to simply ignoring the constraint $\sum_j \mu_j = 1$. Under this augmented model the direct design effect matrix is easy to compute, it is a diagonal matrix with the j diagonal entry

$$\delta_j = \frac{\sum_{U_i\in j-th\ cell} w_i^2}{\sum_{U_i\in j-th\ cell} w_i}. \tag{9}$$

In this case the diagonal entry is also the eigenvalue of the design matrix and therefore the second order approximation for the chi-square testing is easy to compute. Let's call this method the approximate method.

Let's now compare the 3 methods on an LCA example with 6 binary variables and sampling weights $w_i = 1 + exp(u_{i1} + ... + u_{i6})$. Table 1 contains the trace of the design matrix for the H1 model and it is clear from this example that the direct method (6) is the same as the Rao-Thomas method (8). It is also clear from these results that for practical purposes the Approximate method is nearly identical to these methods as well. We have also conducted an extensive simulation study comparing the Rao-Thomas method and the Approximate method and found no essential differences across various models and weights selection.

3 Second Order Correction

The second order correction for the LRT testing is constructed as follows. The LRT and the Pearson test statistics under complex sampling have the asymptotic distribution given in (5). This distribution has mean $Tr(\Delta)$ and variance $2Tr(\Delta^2)$. Let's denote the original chi-square test statistic by X. For example, for the LRT test this will simply be $2(L_1 - L_0)$. We look for a linear combination of this type

$$aX + b \tag{10}$$

which has the same mean and variance as the test statistic under SRS, i.e., with mean d and variance $2d$ where $d = p_1 - p_0$. Simple algebra gives

$$a = \sqrt{\frac{d}{Tr(\Delta^2)}} \tag{11}$$

$$b = d - \sqrt{\frac{d(Tr(\Delta))^2}{Tr(\Delta^2)}}. \tag{12}$$

4 First and Second Order Corrections for the Test of Fit

To summarize we use the following method for obtaining the first and second order adjustments for the test of fit. Let the unadjusted test statistic is X. The first order correction is obtained as

$$\frac{X}{c}$$

where the correction factors c is obtained as

$$c = \frac{Tr(\Delta)}{d} = \frac{Tr(\Delta_1) - Tr(\Delta_0)}{d}$$

where $Tr(\Delta_0)$ is computed by the direct method (6) and $Tr(\Delta_1)$ is computed by the approximate method (15)

$$Tr(\Delta_1) \approx \frac{J-1}{J} \sum_j \delta_j$$

where δ_j are computed by (15). The correction factor $\frac{J-1}{J}$ is used to correct for the augmented cell. In addition, if there are empty cells in the contingency table the correction factor is $\frac{J-1}{J_0}$ where J_0 is the number of non-empty cells in the contingency table, i.e., the number of cells for which δ_j is computable.

The second order adjustment is computed by equations (10-12) where $Tr(\Delta) = Tr(\Delta_1) - Tr(\Delta_0)$. Here $Tr(\Delta_1)$ and $Tr(\Delta_0)$ are computed as in the first order correction described in the previous paragraph. In addition, we use the following approximation to compute $Tr(\Delta^2)$

$$Tr(\Delta^2) \approx \frac{d}{J-1} Tr(\Delta_1^2) \approx \frac{d}{J_0} \sum_j \delta_j^2$$

The first part of this approximation is driven by the fact that the H1 model usually has many more parameters than the H0 model.

5 Simulation Study

In this section we describe the results of a simulation study across 6 different LCA models. We are concerned with the performance of the three chi-square tests of fit: the unadjusted, the first order adjusted and the second order adjusted. We only report The Pearson chi-square test of fit, however the results for the Log-Likelihood chi-square test of fit are very similar. For all 6 models the data is generated according to the same LCA model that we estimate and therefore we expect the test of fit to accept the estimated model, i.e., to lead to a Type I error rate of 5%. In the Tables below we report the average fit statistic, which should match the degrees of freedom d, and the Type I error which should be close to the nominal 5% level. For each model we study the performance of the test statistics at various sample size levels. The data is generated as follows. A large sample of size N is generated. From that sample we select a subsample that will be analyzed according to the following probability selection model. The probability that observation i is included in the subsample is

$$P_i = \frac{1}{1 + Exp(\sum_j \alpha_j U_{ij})}$$

where the vector of parameters α varies across the 6 simulation studies. Varying the α parameters provides different level of informativeness of the sampling weights $w = 1/P$. Because of this data generation the sample size varies across the replications, however that variation is relatively small. We

Table 2: Pearson test of fit average value (type I error) for M1, d=50

N	n	unadjusted	1-st order	2-nd order
1000	116	161(0.98)	78(0.65)	67(0.46)
2500	290	213(1.00)	64(0.63)	59(0.45)
5000	586	238(1.00)	77(0.49)	66(0.36)
10000	1173	231(1.00)	61(0.30)	56(0.13)
20000	2343	226(1.00)	53(0.16)	52(0.04)

report the average sample size n in the tables below as well as the size N of the original sample. The six models and sample selection models we use in this simulation are as follows

- M1. 2-class LCA with 6 binary indicators and $\alpha = (1, 1, 1, 1, 1, 1)$
- M2. 2-class LCA with 6 binary indicators and $\alpha = (0.5, 0.5, 0.5, 0.5, 0.5, 0.5)$
- M3. 2-class LCA with 6 binary indicators and $\alpha = (2, 2, 0.5, 0.5, 0, 0)$
- M4. 3-class LCA with 2 binary and 3 tricotomous indicators and $\alpha = (1, 0, -0.5, 0, 0)$
- M5. 3-class LCA with 10 binary indicators and $\alpha = (1, 0, 0.5, 0, 0, 0, -1, 0, 1.5, 0)$
- M6. 2-class LCA with 10 binary indicators and $\alpha = (1, 0, 0.5, 0, 0, 0, -1, 0, 1.5, 0)$

For each of the models we generate 100 samples and subsamples. These simulation studies are easy to conduct with Mplus and the scripts are available from the authors upon request.

It is clear from these results that in all case the first order corrected test is superior than the unadjusted test and that the second order corrected test is superior than the first order corrected test. For models M1-M4 the second order statistics lead to correct Type I error when the sample size is large, however when the sample size is small Type I error inflation can still be present to some extent. In models M5-M6 where d is large and the H1 model has many empty cells the performance and usefulness of the tests is questionable and should be used very cautiously.

Table 3: Pearson test of fit average value (type I error) for M2, d=50

N	n	unadjusted	1-st order	2-nd order
1000	214	66(0.46)	52(0.08)	52(0.07)
2500	536	66(0.43)	50(0.07)	50(0.04)
5000	1079	66(0.43)	50(0.10)	50(0.08)
10000	2156	66(0.47)	49(0.07)	49(0.04)
20000	4312	66(0.44)	50(0.04)	50(0.03)

Table 4: Pearson test of fit average value (type I error) for M3, d=50

N	n	unadjusted	1-st order	2-nd order
1000	162	158(1.00)	86(0.74)	70(0.47)
2500	406	219(1.00)	89(0.76)	71(0.47)
5000	817	245(1.00)	76(0.60)	65(0.38)
10000	1630	238(1.00)	60(0.33)	56(0.15)
20000	3258	230(1.00)	50(0.10)	50(0.05)

Table 5: Pearson test of fit average value (type I error) for M4, d=81

N	n	unadjusted	1-st order	2-nd order
1000	404	101(0.45)	93(0.19)	91(0.17)
2500	1008	112(0.74)	94(0.32)	92(0.24)
5000	2016	112(0.63)	88(0.22)	87(0.15)
10000	4022	112(0.65)	84(0.17)	84(0.04)
20000	8060	115(0.69)	83(0.13)	83(0.10)

Table 6: Pearson test of fit average value (type I error) for M5, d=991

N	n	unadjusted	1-st order	2-nd order
1000	280	1311(0.92)	1267(0.91)	1208(0.83)
2500	701	1461(1.00)	1353(0.99)	1271(0.98)
5000	1403	1523(1.00)	1345(0.99)	1261(0.99)
10000	2811	1558(1.00)	1296(0.99)	1222(0.98)
20000	5627	1566(1.00)	1217(0.95)	1161(0.92)

Table 7: Pearson test of fit average value (type I error) for M6, d=1002

N	n	unadjusted	1-st order	2-nd order
1000	259	1313(0.93)	1289(0.91)	1225(0.88)
2500	649	1407(1.00)	1350(0.99)	1268(0.99)
5000	1299	1435(1.00)	1329(1.00)	1250(1.00)
10000	2603	1444(1.00)	1275(1.00)	1207(0.97)
20000	5207	1463(1.00)	1214(0.97)	1161(0.90)

6 Stratified and Cluster Sampling

The methods for correcting the Pearson and Chi-Square tests of fit described above easily generalize to the case of stratified cluster sampling. The formulas that are affected by the stratification and the clustering are (7) and (15). Under stratified cluster sampling $Var(L')$ is computed as follows. Let w_{ich} and L_{ich} be the sampling weight and the log-likelihood of individual i in cluster c in stratum h. Then

$$Var(L') = \sum_h \frac{n_h}{n_h - 1} \sum_c (z_{ch} - \overline{z}_h)(z_{ch} - \overline{z}_h)^T \tag{13}$$

where n_h is the number of sampled clusters from stratum h,

$$z_{ch} = \sum_i w_{ich} L'_{ich} \tag{14}$$

is the total score for all individuals in cluster c in stratum h and $\overline{z}_h$ is the average of z_{ch}.

Similarly (15) is computed as follows

$$\delta_j = \frac{\sum_h \frac{n_h}{n_h-1} \sum_c (v_{ch} - \overline{v}_h)^2}{\sum_{U_{ich} \in j-th\ cell} w_{ich}}. \tag{15}$$

where

$$v_{ch} = \sum_{U_{ich} \in j-th\ cell} w_{ich} \tag{16}$$

and $\overline{v}_h$ is the average of v_{ch}.

We now illustrate the performance of the unadjusted, first and second order adjusted tests with a simulation study using complex sampling. We generate data according to a 2-class LCA model with 6 binary indicators. The data is generated according to a two level mixture model. Within each cluster a different LCA model is satisfied and the class membership probabilities vary across clusters. More specifically $log(p_{1k}/(1 - p_{1k})$ and $log(q_1/(1 - q_1)$ are normally distributed cluster level variables. The LCA model is satisfied in the total population as well because the indicator variables are still independent given the class variable. As in the previous section we report the average Pearson test statistic value as well as its Type I error. The number of clusters in the sample is denoted by m and the size of the cluster by n. The results of this simulation are presented in Table 8. It is clear here as well that the adjusted test statistics are more accurate than the unadjusted and the second order adjustment appears to be valuable here as well.

Table 8: Pearson test of fit average value (type I error) under cluster sampling, d=50

m	n	unadjusted	1-st order	2-nd order
100	10	55(0.14)	53(0.05)	53(0.05)
200	20	56(0.10)	49(0.05)	49(0.05)
400	40	68(0.42)	49(0.13)	51(0.11)

7 Conclusion

The adjusted Pearson and Log-likelihood test of fit provide a valuable tool for LCA models estimated with complex samples, while the unadjusted statistics will typically have an inflated Type I error. In some cases however such as large contingency tables or small sample size even the adjusted test of fit have inflated Type I error, however even in these cases the performance of the unadjusted test is much worse than that of the adjusted tests.

References

[1] Asparouhov, T. and Muthen, B. (2005). Multivariate Statistical Modeling with Survey Data. Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.

[2] Asparouhov, T. (2005). Sampling weights in latent variable modeling. Structural Equation Modeling, 12, 411-434. A previous version of this article appeared as Mplus Web Notes: No. 7. Refer to Mplus Papers for the abstract. Paper can be downloaded from here. Click here for more details.

[3] Skinner, C. J. (1989a). Aggregated Analysis: Standard Errors and Significance Tests. In Analysis of Complex Surveys (eds. C.J.Skinner, D.Holt and T.M.F. Smith) 23-58, Wiley.

[4] Skinner, C. J. (1989b). Domain means, regression and multivariate analysis. In Analysis of Complex Surveys (eds. C.J.Skinner, D.Holt and T.M.F. Smith) 59-87, Wiley.

[5] Rao, J.N.K. and Scott, A.J. (1984), On Chi- Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data, Annals of Statistics, 12, 46-60.

[6] Rao, J.N.K. and Thomas, D.R. (1989), Chi-Squared Tests for Contingency Table. In Analysis of Complex Surveys (eds. C.J.Skinner, D.Holt and T.M.F. Smith) 89-114, Wiley.