A parametric bootstrap procedure to perform statistical tests in latent class analysis

Peter van der Heijden, Harm 't Hart and Jos Dessens[1]

Abstract

Latent class analyses of huge data sets are unusual. In this paper we discuss statistical testing problems in this context. Some people try to solve these problems by applying fit indices like the AIC and the BIC in choosing a model. This is incorrect, because these fit indices are based on chi-squared statistics for which the asymptotic behaviour in this context is unknown. We show how the testing problems can be solved by using a parametric bootstrap procedure, also known as a simple Monte Carlo testing procedure. This procedure, proposed earlier by Hope, is worked out in some detail, and applied in the analysis of a large set of variables measuring anti-social behaviour of 2918 youngsters.

1. Introduction

We focus in this paper on latent class analysis of a large data set. The data deal with anti-social behaviour, which is measured by 24 dichotomous variables. Latent class analysis of such huge data sets is unusual. One of the reasons is that latent class analysis defines models for contingency tables, so, if there is a set of k dichotomous items, then a table of 2^k cells is modeled. It will be clear that, in general, only for small k the cells of the contingency table will be reasonably filled. Therefore for large k we cannot rely on asymptotic results, meaning that the usual testing procedures cannot be applied. Furthermore, since the better known computer programs MLLSA (Clogg, 1977) and LCAG (Hagenaars and Luijkx, 1990) require the 2^k frequencies as input, usage of these computer programs for such examples leads to the tedious creation of the correct input file, and computer storage problems.

We know of only one example in the literature where k is much larger than, say, 5. The paper, written by Aitkin, Anderson and Hinde (1981), focuses on a substantive application, namely teaching styles measured by 38 variables, and directs relatively little space to the description of statistical and computational issues. Here we will emphasize these latter issues.

1 Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, Postbus 80.140, 3508 TC Utrecht. A similar version of this paper will appear in a proceedings volume of a workshop on applications of latent class ana latent trait analysis held in Sankelmark, 1994, published by Waxmann Verlag. The comments of anonymous referees are much appreciated.

In latent class analysis there are problems in testing certain hypotheses, and these problems become more severe when the number of cells is much larger than the sample size. In some applications these problems are tackled by using fit indices like the AIC and the BIC, instead of to the usual chi-square statistics. However, since the AIC and the BIC are functions of the chi-squared statistics, they will have the same problems as these statistics. In section 2 we will give a short description of these problems and discuss a solution. In section 3 we then describe a substantive application, and in section 4 we describe the results of this testing procedure in the substantive application. In section 5 we evaluate the testing procedure. Computational issues are described in an appendix.

2. Testing problems in latent class analysis

Usually latent class analysis is fitted by the method of maximum likelihood. Let the 2^k frequencies be indexed by i, then the likelihood, log L, of a particular model A is defined as

$$\log L = \sum_i n_i \log \hat{m}_i^A + \text{constant}$$

where n_i is the observed frequency for cell i, and $\hat{m}_i^A$ is the estimate of the expected frequency under model A. The expression for log L makes clear that it is not necessary to perform calculations for those cells i for which $n_i = 0$. If the number of cells is large compared to the response patterns that are actually observed, then much computer time and space can be saved when a computer program is used that only needs as input data frequencies $n_i > 0$ with the corresponding response patterns. We have written such a program ourselves in APL68000, but other programs that can be used in this way are PANMARK (van de Pol et al, 1991) and LEM (Vermunt, 1993).

If the cells of the contingency table are all reasonably filled, then model A can be tested against the data using the likelihood ratio statistic

$$G_A^2 = 2\sum_i n_i \log \frac{n_i}{\hat{m}_i^A} \qquad (1)$$

as a goodness of fit test where $G_A{}^2$ is asymptotically chi-square distributed with a number of degrees of freedom equal to the number of cells minus the number of independent parameters of model A. In applications, however, two problems may arise.

A first problem that one can encounter in the assessment of the number of latent classes has to do with sparseness of the data: when the number of variables is large, then the cells will not be reasonably filled, and the G^2's are not asymptotically distributed according to specific chi-squared distributions.

In many applications of contingency table analysis the sparseness problem is alleviated by using conditional tests (Agresti, 1990, p.310-313): the G^2-statistic can be partitioned so that it can be used to test differences between models. However, this will not help us in assessing the number of latent classes: if there is a model B in which the number of latent classes is larger than in the true model A,

$$G_{B\cdot A}^2 = 2\sum_i n_i \log \frac{\hat{m}_i^B}{\hat{m}_i^A} \qquad (2)$$

is *not* an asymptotically chi-square distributed variable (with degrees of freedom equal to the difference of the number of independent parameters of model B and A), since certain regularity assumptions are not fulfilled. This problem is well known in the literature on mixture models, which is relevant in this context because the x-class latent class analysis can be considered as a mixture of x independence models, see for example McLachlan and Basford (1988, section 1.10), but is usually not mentioned in introductory texts on latent class analysis. Everitt (1988) shows for some examples the discrepancy of the distribution of G^2 and the chi-squared distribution by means of Monte Carlo studies.

Because assessment of the number of latent classes is important in many applications, we will now discuss a simple Monte Carlo testing procedure that is also known as the parametric bootstrap. This procedure can be used for such an assessment.

Barnard (1963; see also Hope, 1968) originally presented the idea of a simple Monte Carlo test procedure, that works as follows for contingency table problems. Consider a one-sided test of the model (H_0) against the data (the saturated model, H_1), and one is interested in a type-1 error of .05, then

(i). calculate estimates of expected probabilities under H_0 from the sample of size n, and assess the value of the fit criterion under H_0 for the sample

(ii). draw 19 samples of size n from the distribution of estimates of expected probabilities derived in (i), and evaluate the 19 values of the fit criterion,

(iii). if the value of the fit criterion for the sample is larger than the 19 values found in (ii), then reject H_0, because this value is then the 20th value, corresponding with the type-1 error of .05.

The 19 values in step ii. constitute a Monte Carlo distribution of the fit criterion. The idea to reject H_0 if the sample value is larger than the highest value of this Monte Carlo distribution is the usual idea underlying tests, namely, in this case it is unlikely that the model has generated the sample (although the type-1 error of such a decision is .05). This procedure is easily extended to two-sided tests, and Hope (1968) showed that the power of this test procedure increases when the number of Monte Carlo samples increases.

This procedure is also easily extended to conditional tests. Let H_0 define a model that is a special case of the restrictive model under H_1. Let there be a fit criterion that can be used to measure the difference between the models under H_0 and H_1. Then

(i). calculate estimates of expected probabilities under H_0 and under H_1 from the sample of size n, and assess the value of the fit criterion for the difference between H_0 and H_1 for the sample

(ii). draw 19 samples of size n from the distribution of estimates of expected probabilities derived in (i) under H_0, and evaluate the 19 values of the fit criterion for the difference between H_0 and H_1,

(iii). if the value of the fit criterion for the sample is larger than the 19 values found in (ii), then reject H_0 in the conditional test procedure.

Readers well aware of bootstrap methodology (see, for an introduction, Efron and Tibshirani, 1993), will recognise this procedure as an example of the parametric bootstrap, i.e. in the well-known non-parametric bootstrap bootstrap samples are drawn from the data, whereas in the parametric bootstrap samples are drawn from a distribution derived from a sample by imposing a restrictive model on the sample.

To our knowledge, the only application of this parametric bootstrap procedure to latent class analysis is by Aitkin et al. (1981), who used the procedure for unconditional tests only. The conditional procedure is used by de Soete and Winsberg (1993) to assess the number of dimensions of the latent class vector model. Bollen and Stine (1992) employ a similar procedure in the context of structural equation models.

In this paper we will use only the procedure for unconditional tests. Numerical and computational issues are discussed in an appendix. The APL-computer programs are available from the first author upon request. There also exists a beta version of a new PANMARK release with an option to do parametric bootstraps. This program is certainly more userfriendly than our program.We refer to Langeheine, Pannekoek and van der Pol (in press) for related work in this area.

3. The example: anti-social behaviour

The data stem from the first wave of a longitudinal study into the life course of youngsters (see Meeus and 't Hart, 1993). There are 2918 youngsters, aged between 12 and 24. They indicated on a long list of anti-social behaviours whether or not they performed these behaviours in the year preceding the interview. We started by studying the set of those questions to which at least one youngster answered 'yes'. The questions are presented in table 1. The research question is how many typical groups of youngsters there are, in terms of anti-social behaviour. Since it is well-known from criminological research that the prevalence of anti-social behavior is related to age and sex, we also study what role these variables play in the explanation of anti-social behaviour. We try to answer these questions using latent class analysis. In the context of anti-social behavior, earlier examples of latent class analysis are reported by Brownfield and Sorensen (1989), Sorensen and Brownfield (1989) and Fergusson et al. (1993).

Table 1: the variables measuring anti-social behaviour, and their last year prevalence in a sample of 2918 Dutch youngsters, aged between 12 and 24. The questions a to f were not used in the LCA that was conducted, because the number of local maxima that occurred was relatively large. The LCA of the variables 1 to 18 is therefore overemphasising the relatively less anti-social behaviours. The 18 variables are ordered according to their last year prevalence, this order not being the same as the order in the questionnaire.

You filled in that you ever ..... Did you do this last year? yes/no.	
a. Stealing a car?	1
b. Pick pocketing?	2
c. Beating up a family member	2
d. Threatening with a weapon or beating up somebody to obtain money or goods	5
e. Wounding with a knife or other weapon	7
f. Stealing from a car	11
1. Raising a fire	34
2. Stealing from a telephone booth or slot machine	36
3. Selling stolen goods	37
4. Beating up a non-family member	39
5. Stealing something else	44
6. Stealing a bicycle, moped or motor	57
7. Committing burglary, sneaking in	68
8. Graffiti	89
9. Stealing from school children	106
10. Buying stolen goods	127
11. Stealing from a shop	138
12. Stealing from home	162
13. Stealing from work	169
14. Dodging fare (train)	204
15. Being involved in a fight	245
16. Vandalism	343
17. Dodging fare (bus, trolley, subway)	407
18. Having a knife or another weapon	422

4. Results

In table 1 the 24 items were described, and our original aim was to include all 24 items in an analysis. However, for technical reasons we failed in this respect. We started off by performing latent class analyses on 24 variables, and it turned out that the likelihood function had many local optima for the models with three or more latent classes. We therefore excluded items a, b and c in table 1, having last year prevalence 1, 2 and 2. Subsequently we performed latent class analyses of the remaining 21 variables. Here we had the same problems for solutions with three or more latent classes, so we excluded variables d, e and f in table 1, having a last year prevalence of 5, 7 and 11. Thus we ended up with an analysis of 18 variables. Because six extreme forms of anti-social behaviour were excluded, the interpretation of the results of the analyses will necessarily be in terms of relatively more common types of anti-social behaviour.

In Table 2, the Monte Carlo distributions of the G^2 - statistic is shown, which can be compared with the G^2-value that is obtained in our sample of 2918 youngsters. Monte Carlo distributions are derived from 50 parametric bootstrap samples. For each analysis the Monte Carlo G^2-values are ordered according to their value, and in order to give a rough idea about their distribution, specific points of the distribution are given.

We started with ordinary LCA, ignoring the possible grouping of the 2918 youngsters by age and gender. Table 2a shows that the latent class models with one, two or three latent classes clearly have to be rejected, because the sample G^2-value is higher than the Monte Carlo distribution, indicating that it is very unlikely that these models have generated the data. Given the parametric bootstrap procedure, we cannot reject the model with four latent classes.

We found this model to be less parsimonious in terms of the number of latent classes, than we hoped for. A solution with only two latent classes, for instance, would simply imply that anti-social behaviour is one-dimensional, and there is no specialisation in deviating. There are criminological theories supporting this view (for example, see Gottfredson and Hirshi, 1990). Therefore we turned to a less restrictive model, namely simultaneous LC model with homogeneous conditional probabilities but heterogeneous class sizes. By controling for age and gender in this way, the number of latent classes necessary to describe the data could hopefully be reduced. Under this model, groups are easily compared, because the only parameters needed to compare groups are the class sizes. We stopped investigating this model for the solution with 5 latent classes. It is unclear whether this model should be rejected or not, because the sample G^2 of 3590 falls somewhere around $.04<p<.06$. It should be emphasised that the number of 50 parametric bootstrap samples is much too small to estimate the end of the Monte Carlo distribution with high precision. However, our aim to simplify the interpretation failed, because the number of latent classes went up instead of down. (We want to emphasise here that the ordinary latent class model is not equivalent to the simultaneous latent class model with homogeneous class sizes and homogeneous conditional probabilities, because the former model is a model for a table of size 2^{18}, whereas the latter

Table 2. Test results. Presented are the G^2 - test values for the analysis of 18 variables. Monte Carlo distributions are derived from 50 parametric bootstrap samples. For each analysis the Monte Carlo G^2-values are ordered according to their value, and in order to give a rough idea about their distribution the first, second, fifth, twentieth, thirtieth, forty-fifth, forty-nineth and fiftieth values are given. For example, the second column shows the solution for one latent class (LC=1), 18 independent parameters are fitted, the G^2-value for the sample is 3958, the smallest (i.e. the first) value from the Monte Carlo distribution is 671 and the largest is 875, showing that if the model with 1 latent class would be true, then a sample resulting in G^2 = 3958 is so unlikely that we reject this model to be true.

Table 2a. Test results for LCA. The Monte Carlo distribution for 5 latent classes is not derived because the model with 4 latent classes could not be rejected.

	LC=1	LC=2	LC=3	LC=4	LC=5
#par	18	37	56	75	94
G^2	3958	2358	2150	1944	1880
	$p<.02$	$p<.02$	$p<.02$	$.1<p<.4$	
1	671	1733	1581	1615	
2	696	1759	1602	1634	
5	707	1802	1722	1667	
20	757	1935	1798	1763	
30	807	1997	1856	1833	
45	829	2115	1967	1951	
49	853	2148	1996	1982	
50	875	2184	2001	2016	

Table 2b. Test results for simultaneous LCA with heterogeneous class sizes and homogeneous conditional probabilities. The Monte Carlo distribution for 6 latent classes is not derived because the model with 5 latent classes could not be rejected, and the distribution for one latent class is not derived because it was clear in advance that it would clearly be rejected.

	LC=1	LC=2	LC=3	LC=4	LC=5	LC=6
G^2	6266	4438	4041	3772	3590	3469
#par	18	44	70	96	122	
	$p<.02$	$p<.02$	$p<.02$	$p<.02$	$.04<p<.06$	
1		3416	3244	3108	3032	
2		3475	3270	3138	3087	
5		3637	3321	3235	3131	
20		3742	3507	3336	3244	
30		3834	3580	3440	3326	
45		3987	3737	3581	3531	
49		4077	3817	3724	3599	
50		4084	3842	3749	3669	

Table 2, continued

Table 2c. Test results for simultaneous LCA with heterogeneous class sizes and heterogeneous conditional probabilities. In the first columns we find the results for the latent class model with one and with two classes. The solution with two latent classes fits remarkably well: .6<p<.9. On the right we have split up this G^2 of 3418 over the eight age-gender groups, and for each of the groups the fit is adequate: no p-value is smaller than .05.

	LC=1	LC=2
G^2	4972	3418
#par	144	296
	p<.02	.6<p<.9
1	1757	3200
2	1781	3289
5	1826	3325
20	1930	3552
30	1969	3708
45	2053	3998
49	2086	4144
50	2104	4275

	LC=2	
group 1	396	.30<p<.32
2	976	.18<p<.20
3	623	.26<p<.28
4	358	.88<p<.90
5	217	.32<p<.34
6	426	.24<p<.26
7	251	.28<p<.30
8	171	.16<p<.18
	3418	

model is a model for a table of size 8 x 2^{18}. Therefore the chi-squares of table 2a cannot be compared with the chi-squares of table 2b.)

Therefore we turned to the simultaneous LC model with both heterogeneous class sizes and heterogeneous conditional probabilities. This model with two latent classes fitted very well for each of the eight age-gender groups (see table 2c). Although this leads to an enormous increase of independent parameters fitted (namely from 139 for the 5 latent class model in table 2b to 303 independent parameters for the 2 latent class model in table 2c), the interpretation is considerably simplified, because we can now say that for each combination of age and gender anti-social behaviour is one-dimensional, but the way in which anti-social youngsters express anti-social behaviour depends on age and gender. This interpretation says that , although the class structures are different , the latent class structures are conceptually identical. In principle it is also possible to argue that these 8 latent class structures measure something which is also conceptually different, but this does not seem likely given the interpretation of both classes in each structure, and the class size probabilities, that follow the usual age-peak (see below).

We now interpret the parameter estimates of the partly heterogeneous simultaneous latent class model with five latent classes in table 2b and the completely heterogeneous simultaneous latent class model with two latent classes in table 2c. In table 3 we find the solution for simultaneous LCA with five latent classes, with heterogeneous class size estimates at the top and homogeneous conditional probabilities at the bottom. The

Table 3: Solution for simultaneous LCA with heterogeneous class sizes and homogeneous conditional probabilities, model with 5 latent classes. The latent class probabilities add up to 1 rowwise. For the conditional probabilities only the estimates for the probabilities to commit the anti-social behaviour is reported. For example, for variable 1, given that one falls into latent class 1, the estimated probability to say 'yes' is .000, and hence the estimated probability to say 'no' is 1.000. Over all latent classes, the (weighted) average probability to say 'yes' is equal to the observed proportion of saying 'yes', which is .012 for variable 1.

heterogeneous class sizes

	Latent classes				
	1.	2.	3.	4.	5.
males					
12-14	.450	.321	.178	.000	.050
15-17	.333	.339	.042	.145	.141
18-20	.422	.191	.000	.313	.074
21-24	.622	.000	.000	.336	.042
females					
12-14	.666	.000	.328	.000	.006
15-17	.595	.000	.238	.141	.026
18-20	.760	.009	.023	.187	.021
21-24	.826	.000	.004	.165	.005

homogeneous conditional probabilities

	Latent classes					
	1.	2.	3.	4.	5.	observed proportion
variables						
1	.000	.035	.005	.000	.160	.012
2	.002	.018	.037	.005	.097	.012
3	.000	.008	.000	.020	.186	.013
4	.000	.062	.000	.007	.124	.013
5	.000	.012	.088	.012	.063	.015
6	.003	.000	.000	.045	.233	.020
7	.004	.045	.051	.022	.165	.023
8	.001	.026	.134	.008	.264	.031
9	.000	.089	.080	.064	.185	.036
10	.005	.014	.027	.133	.333	.044
11	.005	.043	.115	.050	.437	.047
12	.006	.158	.243	.023	.148	.056
13	.022	.027	.003	.194	.236	.058
14	.004	.000	.034	.251	.514	.070
15	.012	.244	.073	.125	.509	.084
16	.017	.307	.367	.064	.603	.118
17	.020	.058	.182	.452	.663	.139
18	.038	.535	.071	.188	.630	.145

homogeneous conditional probabilities are more easily interpreted by comparing them with the observed proportions of the various forms of anti-social behaviour. This shows that the first latent class is clearly the non-anti-social class, because all estimates are much lower than the observed proportions. The heterogeneous class size estimates show that *not* being in this class is peaked at age 15-17, both for boys and girls. Such an age-peak in anti-social behaviour is well known from the literature. Girls tend to show much less anti-social behaviour than boys (i.e. they fall more often into this class). In the fifth latent class all the estimates to perform anti-social behaviour are much higher than is observed for the whole group. For both boys and girls this peaks at age 15-17, but much more boys than girls fall into this class. We coin this a multi-anti-social class. The second latent class is used predominantly by males 12-20. Variables 1, 2, 4, 7, 9, 12, 15, 16 and 18 have estimates much higher than their corresponding observed proportions. These are items that are referring to showing off, vandalism, violence and stealing, and it seems that items having expressive anti-social behaviours against people are overrepresented in this class. The third latent class is used predominantly by the younger age groups, but especially by females. Variables 2, 5, 7, 8, 9, 11, 12, 16 and 17 have estimates much higher than their corresponding observed proportions. These items mainly have to do with stealing and vandalism, and it seems that items having expressive anti-social behaviours against objects are overrepresented in this class. The fourth latent class is used predominantly by older age groups, especially by boys. Variables 3, 6, 9, 10, 13, 14, 15, 17 and 18 have estimates much higher than their corresponding observed proportions. These variables show that this kind of anti-social behaviour is mainly of an instrumental type. It is not completely straightforward to interpret latent classes 2, 3 and 4, as we did above. It is also possible, of course, to think of these classes as giving descriptions of what kind of behaviour is in groups of a particular age and gender, more often exhibited than on average.

Table 4 shows the results for simultaneous LCA with two latent classes, having both heterogeneous class sizes as well as heterogeneous conditional probabilities. Tentatively, we interpret this solution as an indication that for each age-gender combination that there are two groups of youngsters: anti-social ones and non-anti-social ones. This is in line with the class sizes of the anti-social behaviours, which peak both for males and for females at age 15-17. The conditional probabilities then show which behaviours are more likely to be picked by people from a specific age-gender combination. It is enlightening to study the estimates to perform anti-social for each anti-social behaviour separately, because it shows when particular behaviour peaks as a function of age. For example, "graffiti" (variable 8) peaks for boys at age 15-17, but for girls at age 12-14, going down when they get older.

5. Discussion

From a substantive point of view, our results support criminological theories stating that there is no specialization in crime (see, for example, Gottfredson and Hirshi, 1990). In this discussion we focus on the statistical and computational aspects of this study. There are a number of points that need special attention.

Table 4: parameter estimates for the simultaneous latent class model with two latent classes, heterogeneous class sizes and heterogeneous conditional probabilities. First four columns are estimates to answer 'no' for males of four age groups, columns 5 to 8 are estimates to answer 'no' for males of four age groups, columns 9 to 12 are estimates to answer 'no' for females of four age groups, columns 13 to 16 are estimates to answer 'no' for females of four age groups. '---' indicates that in this particular age-sex group none of the respondents answered 'yes to this particular variable.

Sex	males								females							
Class	1				2				1				2			
Age	12-14	15-17	18-20	21-24	12-14	15-17	18-20	21-24	12-14	15-17	18-20	21-24	12-14	15-17	18-20	21-24
Class sizes	.882	.768	.816	.829	.118	.232	.184	.171	.887	.879	.883	.905	.113	.121	.117	.095
Conditional probabilities																
Variable																
1	.000	.020	.003	.000	.164	.124	.069	.020	.000	---	.000	.000	.056	---	.023	.026
2	.007	.007	.005	.000	.059	.062	.024	.020	.011	.009	.003	.002	.052	.115	.023	.055
3	.000	.013	.000	.000	.082	.125	.176	.119	---	.000	---	---	---	.018	---	---
4	.013	.011	.014	.000	.124	.101	.131	.040	---	.000	.000	.003	---	.018	.045	.000
5	.029	.007	.003	.004	.000	.083	.017	.020	.028	.000	.000	.000	.198	.053	.045	.052
6	.000	.000	.028	.021	.055	.190	.148	.096	---	.000	.005	.008	---	.053	.026	.028
7	.007	.017	.018	.020	.247	.156	.031	.062	.020	.017	.000	.000	.043	.071	.068	.026
8	.027	.002	.000	.004	.153	.215	.144	.060	.005	.010	.003	.000	.441	.302	.046	.026
9	.029	.046	.038	.000	.168	.165	.120	.060	.008	.022	.005	.007	.163	.144	.207	.063
10	.000	.017	.026	.036	.109	.196	.301	.379	.000	.000	.020	.023	.140	.213	.162	.151
11	.030	.023	.016	.000	.271	.294	.199	.159	.000	025	.013	.011	.309	223	.242	.185
12	.101	.083	.025	.004	.154	.177	.083	.060	.062	.039	.013	.000	.241	.339	.012	.104
13	---	.027	.074	.117	---	.154	.373	.206	.004	.016	.027	.046	.000	.114	.228	.344
14	.004	.037	.035	.007	.242	.372	.324	.582	.000	.049	.011	.018	.028	.299	.528	.379
15	.061	.092	.085	.044	.528	.432	.486	.343	.020	.027	.031	.011	.122	.282	.106	.028
16	.147	.123	.081	.028	.653	.542	.345	.201	.078	.075	.027	.011	.653	.446	.273	.055
17	.035	.093	.120	.034	.339	.525	.602	.849	.047	.121	.058	.021	.335	.449	.663	.709
18	.199	.237	.162	.114	.621	.638	.510	.317	.026	.033	.045	.041	.192	.258	.310	.079

1. Local maxima. As discussed at the beginning of section 4, we found that local maxima made it sometimes difficult to find a final solution. Our impression is that there may not be too many variables with a proportion "yes" or "no" close to zero, because this leads to problems in the estimation of the model. There are two problems, and an assessment of the importance of both problems needs a rigorous investigation that falls outside the scope of this paper. One problem is that, in the EM algorithm, once an estimate has value 0 in the iteration process, it cannot leave that value. This problem is particularly eminent when one or more variables are skewed. This shows that, although the likelihood is well defined even if many observed frequencies are zero, in practice it can result into estimation problems. A second problem is that, in the general case, the likelihood function in LCA can have multiple maxima. Our impression is that the number of maxima increases when the data that have to be spread out over the latent classes become sparser. So this phenomenon is particularly a problem when the number of latent classes increases. See also Aitkin at al. (1981, p.425-427).

2. Number of bootstrap samples. In our study we have chosen a number of 50 bootstrap samples for each test. This number seems to be far too low to have a good estimate of the Monte Carlo distribution of the G^2-statistic. It is unclear to us how large this number should be. Hope (1968) proves that the power of the test rises if the number of bootstrap samples rises.

3. Which statistic to use? Hope (1968) emphasises that the statistic used for the assessment of the fit of the model should be adequate for this purpose. Although G^2 is useful in those circumstances where its asymptotic behaviour is well known (i.e. if the sample size goes to infinity it follows a chi-squared distribution if H_0 is true), it is not clear whether this statistic is useful when the assumption of an infinite sample size is not approximated at all. In circumstances like this Hope (1968) advises to use more than one statistic, and hopefully all these statistics lead to the same conclusion regarding H_0. Possible test statistics are Pearson's chi-square that sums (observed - expected)^2/expected over all cells. However, for our example this would mean that we would have to calculate the expected frequency for all cells in the contingency table, and this would mean a computationally rather expensive task. In other examples it would certainly be advisable to use this statistic as well, especially because of the unclarity about the usefulness of G^2 versus Pearson's chi-square in the light of small sample size. Other statistics are AIC, BIC and CAIC, but these would all lead to the same conclusion as bootstrapping tests using G^2, because they are linear transformations of this statistic. What seems to be needed is to show by simulation studies that this procedure works well, but this involves a lot more work. A comparable simulation study is performed by de Soete and Winsberg (1993) to show that this procedure worked in the context of the latent class vector model.

4. Parameter estimates equal to zero for the sample. A referee has pointed out that succes of our bootstrap procedure is not warranted because some of the parameter estimates for the sample are on the boundary of the parameter space, i.e. they are equal to zero or one. "Under resampling from this population, these parameters will show no variation at all. However, in the true population these parameters may be close to the boundary, but not on it. Sampling from the true population will result in some variability for these estimates. If this reasoning is correct, the two sampling distribution will not be the same and the validity of the bootstrapping will be questionable." Although we agree with this viewpoint, nothing is known to what degree the bootstrapping results will be invalid in these cases. In general we expect minor effects. The underpinning of this expectation, however, requires further research in that we will be able to assess to what extent zero parameter estimates will invalidate the bootstrapping procedure.

References

Aitkin, M., Anderson, D., and Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, Series A, 144*, 419-461.

Barnard, G.A. (1963). "Contribution to the discussion of a paper by M.S. Bartlett". *Journal of the Royal Statistical Society, Series B, 25*, 294.

Bollen, K.A. and Stine, R.A. (1992). Bootstrapping goodness-of-fit measures in the context of structural equation models. *Sociological methods and research, 21*, 205-229.

Brownfield, D. and A.M. Sorenson (1989) A Latent Strucure Analysis of Delinquency. *Journal of Quantitative Criminology, 3*: 103-124.

Clogg, C.C. (1977). *Unrestricted and restricted maximum likelihood latent structure analysis: a manual for users*. Pennsylvania State University, Population Issues Research Office, Working paper No. 1977-9.

Davis, C.S. (1993). The computer generation of multinomial random variates. *Computational statistics and data analysis, 16*, 205-217

de Soete, G. and S. Winsberg (1993). A latent class vector model for preference ratings. *Journal of classification, 10*, 195-218.

Efron, B. and R.J. Tibshirani (1993). *An introduction to the bootstrap*. New York: Chapman Hall.

Everitt, B.S. (1988). A Monte Carlo investigation of the likelihood ratio test for number of classes in latent class analysis. *Multivariate Behavioral Research*, 23, 531-538.

Fergusson, D.M., L.J. Horwood and M.T.Lynskey (1993). *The comorbidities of adolescent problem behaviour: a latent class model.* Unpublished research note, Department of Psychological Medicine, Christchurch, New Zealand.

Gottfredson, M.R. and T. Hirshi (1990). *A general theory of crime*. Stanford: Stanford University Press.

Hagenaars, J.A. and Luijkx, R. (1990). *LCAG. A program for latent class models and other loglinear models with latent variables with and without missing data.* Tilburg: Tilburg University, Department of Sociology, Working Paper Series #17.

Hope, A.C.A. (1968). A simplified Monte Carlo significance procedure. *Journal of the Royal Statistical Society, Series B, 30*, 582-598.

Kemp, C.D. and A.W. Kemp (1987). Rapid generation of frequency tables. *Applied Statistics, 36*, 277-282

Langeheine, R., J. Pannekoek and F. van der Pol. (in press). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological methods and research.*

McLachlan, G.J. and Basford, K.E. (1988). *Mixture models.* New York: Dekker.

Meeus, W. en H. 't Hart (1993). *Jongeren in Nederland*. Amersfoort: Academische Uitgeverij.

Sorenson, A.M. and D. Brownfield (1989). Patterns of Drug Use: Inferences from Latent Structure Analysis. *Social Science Research, 18:* 271-290.

van der Pol, F. Langeheine, R. and de Jong, W. (1991). PANMARK user manual; PANel analysis using MARKov chains, version 2.2. Voorburg: Netherlands Central Bureau of Statistics.

Vermunt, J. (1993) LEM: Log-linear and event history analysis with missing data using the EM algorithm. Tilburg: WORC PAPER 93.09.015/7.

Appendix

Although the Monte Carlo test procedure is remarkably simple in its presentation, there are some details that have to be kept in mind while using it in the context of latent class analysis of a large number of variables.

We now describe phase (ii) of the unconditional test of H_0 described in section 2. Let there be k dichotomous variables, so that there are, theoretically, 2^k response patterns. The estimates of the parameters found in step (i) define a multinomial distribution over the 2^k cells of the contingency table.

Kemp and Kemp (1987) and Davis (1993) have given useful advises on how to draw a sample from a multinomial distribution. A way that is simple to program, though computationally not the most efficient one, is the so-called direct or naive method. First the 2^k cumulative estimated probabilities are calculated. Then a random observation is drawn from a uniform distribution between 0 and 1, and this observation is assigned to a cell using the cumulative probabilities. For a sample of size n, n such random observations are drawn, and assigned to the 2^k cells. This then provides the sample drawn from the multinomial. A simple modification, that is particularly important for our application, is first drawing the n random observations, then ordering the multinomial probabilities according to their size, so that hopefully not the complete cumulative distribution needs to be calculated (see below).

To avoid calculating all the 2^k probabilities, as well as calculating all cumulative probabilities, we take the following steps:

1. draw n=2918 observations from a uniform distribution between 1 and 10^{15} (the maximum floating point precision of APL68000) and order them. We go through the 2^{18} response patterns from 1...1111, via 1...1112, 1...1121, etcetera to 2...2222. However, by first ordering the k=18 variables from lowest proportion "yes" to highest proportion "yes", it turns out that the latter probabilities to be calculated are on average smaller than the first former ones (this is a result from the fact that the observed one-way marginal probabilities of the table of order 2^{18} are equal to the corresponding margins of estimates of expected probabilities). Thus we avoid calculations for as many response patterns as possible. For each response pattern
2. calculate the probability of a particular response pattern, multiply that probability with 10^{15}, derive the cumulative probability, and see if there are any random observations to be assigned to this response pattern. If so, store the response patterns with the number of random observations assigned to it, and then go to the next response pattern, if not, go directly to the next response pattern. A possible problem here is that there are estimated probabilities so small (much smaller than 10^{-15}) that it is not possible to assign any random observations to them. We can only hope that this numerical problem does not influence our results a lot, which is likely because the probability that any of the 2918 random observations is assigned to such a cell has a very small probability indeed.
3. end if all the 2918 random observations are assigned. Thus the first parametric bootstrap sample is obtained.
4. perform an LCA, using the parameter estimates from the original sample as starting values. If the G^2 of the bootstrap sample turns out to be larger than the G^2 of the original sample it is checked whether the bootstrap sample G^2 is actually the G^2 of a local maximum by doing three analyses with random starting values. If the G^2 of the bootstrap is smaller, we do not have to check for local maxima. The reason for this is that the only aspect of the Monte Carlo distribution we use is the number of bootstrap samples having a larger G^2 than the original sample.

Calculating a probability of a response pattern from the parameter estimates takes a rather large part of all the computing time. Therefore in phase 1 we draw 50 strings of 2918 random observations, in phase 2 we calculate the probability and cumulative probability once and assign observations from each of the 50 strings to 50 bootstrap samples.

For simultaneous latent class analysis we have simply performed the above routine 8 times, once for every group. In doing this, we have conditioned on the sample size of each of the groups. So this yields 8 Monte Carlo distributions of G^2 (see the right part of table 2c). We have created the Monte Carlo distribution of the overall model (i.e. the 8 groups simultaneously) by adding the eight distributions where the 50 values of each distribution were placed in a random order (see column 2 of table 2c).

Ontvangen: 20 mei 1995
Geaccepteerd: 7 februari 1996