Sociological Methods & Research
Volume 36 Number 1
August 2007 87-111

10.1177/0049124107301965
http://smr.sagepub.com
hosted at
http://online.sagepub.com

Latent Class Analysis With Sampling Weights

A Maximum-Likelihood Approach

Jeroen K. Vermunt
Tilburg University, Netherlands
Jay Magidson
Statistical Innovations, Inc., Boston, MA

The authors illustrate how to perform maximum-likelihood estimation in latent class (LC) analysis when there are sampling weights. The methods are natural extensions of the approaches proposed by Clogg and Eliason (1987) and Magidson (1987) for dealing with sampling weights in the log-linear analysis of frequency tables. For the log-linear form of the LC model, the approach corresponds to a special case of Haberman's (1979) log-linear LC model with cell weights. This approach can also be applied to the probability formulation of the LC model with cell weights, which can accommodate many indicators. The authors propose an efficient estimation-maximization algorithm for estimating the parameters for this formulation. A small simulation study shows that the probability estimates obtained by this approach compare favorably to other weighting approaches. Several empirical examples are provided to illustrate various possible weighting methods in LC analysis.

Keywords: *latent class analysis; mixture model; complex sampling; log-linear analysis; EM algorithm*

Data sets from surveys often contain case or sampling weights. Such weights can be used to adjust for cases that are under- or overrepresented in the sample because of the sampling scheme. The prevailing method for dealing with sampling weights in the analysis of frequency tables is to construct a weighted observed frequency table and then to analyze it as if it were an unweighted table. This approach is referred to as the pseudo-maximum-likelihood (ML) estimation method (Rao and Thomas

Authors' Note: Please address correspondence to Jeroen K. Vermunt, Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, 5000LE Tilburg, Netherlands; e-mail: J.K.Vermunt@uvt.nl.

Table 1
Cross-Tabulation of Three Dichotomous Indicators From a Hypothetical Survey Containing Sampling Weights

j	Y_1	Y_2	Y_3	n_j	$n_j^{(w)}$	$n_j^{(w)}/n_j$	z_j
1	1	1	1	176	153.25	0.87	1.15
2	1	1	2	45	56.18	1.25	0.80
3	1	2	1	54	63.39	1.17	0.85
4	1	2	2	111	114.56	1.03	0.97
5	2	1	1	76	75.05	0.99	1.01
6	2	1	2	116	126.78	1.09	0.91
7	2	2	1	154	141.90	0.92	1.09
8	2	2	2	399	431.89	1.08	0.92

1988; Skinner, Holt, and Smith 1989). An alternative method is to ignore sampling weights in parameter estimation, possibly combined with a second step in which certain model parameters are corrected for weighting. In the context of latent class (LC) analysis and mixture modeling, the pseudo-ML approach has been advocated by Patterson, Dayton, and Graubard (2002) and Wedel, Ter Hofstede, and Steenkamp (1998) and the two-step approach by Vermunt (2002) in a commentary on the Patterson et al. (2002) article. Both approaches are implemented in the LC software package Latent GOLD (Vermunt and Magidson 2005).

To illustrate the various available options for dealing with sampling weights in LC analysis, consider the three-way cross-tabulation displayed in Table 1, which contains information on three dichotomous indicators obtained from a hypothetical survey containing sampling weights. Suppose we wish to construct a two-class LC model based on these three indicators denoted by Y_1, Y_2, and Y_3. More precisely, we wish to estimate the proportion of persons belonging to LC 1 and possibly to compare this proportion across years or subgroups. The two current options are to use either the unweighted observed frequencies (n_j) or the weighted observed frequencies ($n_j^{(w)}$) in the LC analysis. The latter approach yields pseudo-ML estimates for the LC model parameters. When using the unweighted frequencies, we may correct the sizes of LCs after obtaining estimates of the model parameters; that is, we correct the unconditional class membership probabilities for the fact that sampling weights may be correlated with LC membership.

In this article, we describe a third approach to weighting in LC analysis that extends the approach proposed by Clogg and Eliason (1987) and Magidson (1987) for ML estimation with sampling weights in log-linear analysis, in which the unweighted observed frequencies (n_j) are used as data and the inverse of the cell-specific weights (z_j) are included as a term in the log-linear model (see also Agresti 2002:391). Our generalization of this ML approach to LC models turns out to be a special case of Haberman's (1979, 1988) log-linear model with cell weights for frequency tables derived by indirect observations. A drawback of the estimation procedure proposed by Haberman for log-linear LC models is that it can be used only when the number of indicators is small. To be able to use the weighting method also with large numbers of indicators, we adapt the weighting method to the probability formulation of the LC model and develop an efficient EM algorithm for parameter estimation. Although here we will concentrate on LC models, the resulting weighting method is applicable to a much broader class of models for frequency tables.

It should be noted that pseudo-ML estimation and accompanying methods for computation of design-corrected standard errors (jackknife, linearization, bootstrap, etc.) can deal not only with sampling weights, but also with clustering and stratification, which are other relevant aspects of complex sampling designs. A problem associated with pseudo-ML estimation is, however, that standard goodness-of-fit tests and related measures such as Akaike information criterion and Bayesian information criterion (BIC) can no longer be used. This problem does not occur when using the ML approach described in this article. In other words, the proposed ML approach, while more limited than pseudo-ML estimation because it can deal with only one aspect, sampling weights, of the sampling design, has the advantage that it does not complicate model testing.

A question of interest is whether the new ML estimation method for dealing with sampling weights in LC analysis yields more reliable estimates of the population parameters of interest than the commonly used methods. To answer this question, we compare the various approaches in a small simulation study. Since it can be expected that the performance of the weighting methods depends on the homogeneity of the measurement model (Vermunt 2002), special attention is paid to this issue in the design of the simulation study.

This report is organized as follows. The next section describes log-linear analysis with sampling weights. Then, we show how to extend this method to the log-linear and probability formulations of the LC model.

Subsequently, we report the results from a simulation study and illustrate the various weighting methods using several empirical examples. We end with a short discussion.

Sampling Weights in Log-Linear Analysis

Let w_i denote the sampling weight for case i, and let δ_{ij} be an indicator variable taking on the value 1 if case i falls into cell j of the frequency table and 0 otherwise. The unweighted observed frequency for cell j, n_j can be obtained as follows:

$$n_j = \sum_i \delta_{ij}.$$

In addition to this unweighted cell frequency, we can define the weighted observed frequency, $n_j^{(w)}$, computed as

$$n_j^{(w)} = \sum_i \delta_{ij} w_i.$$

Typically, a weighted analysis is performed by using the weighted observed frequencies $n_j^{(w)}$ as data in an "unweighted" analysis, which amounts to using pseudo-ML estimation (Skinner et al. 1989). Clogg and Eliason (1987) and Magidson (1987) showed that by correcting for the fact that certain cases are over- or underrepresented in the sample, this procedure gives the correct parameter estimates for saturated log-linear models. However, such weighting distorts the data so that certain assumption underlying the goodness-of-fit tests for nonsaturated models will no longer be valid. Specifically, if the Pearson and likelihood-ratio goodness-of-fit tests are computed by comparing the estimated frequencies with the weighted observed frequencies $n_j^{(w)}$, these statistics will no longer be distributed as χ^2 because $n_j^{(w)}$ does not represent the number of independent observations with answer pattern j. Only n_j, the unweighted observed frequency, represents the number of independent observations.

Haberman (1978) proposed a log-linear model with cell weights that is defined as

$$m_j = \exp(\mathbf{x}_j \boldsymbol{\beta})\ z_j = h_j\, z_j. \tag{1}$$

Here, m_j is an unweighted expected cell frequency, h_j is a weighted expected cell entry, z_j is a cell weight, and $\mathbf{x}_j\boldsymbol{\beta}$ represents the linear term of the log-linear model.

Clogg and Eliason (1987) and Magidson (1987) showed that this log-linear model can be used for ML estimation with sampling weights. This requires equating z_j to the inverse of the aggregated sampling weight for cell j. More precisely,

$$z_j = \frac{n_j}{n_j^{(w)}}.$$

By rewriting the model defined in equation (1) as

$$h_j = m_j \frac{n_j^{(w)}}{n_j} = \frac{m_j}{z_j},$$

it is easier to see that the weighted expected cell count (h_j) equals the unweighted expected cell count (m_j) times a cell-specific weight (z_j^{-1}). Goodness-of-fit tests will now be valid because these are based on the standard comparison of m_j and n_j (Hendrikx 2002).

Sampling Weights in Log-Linear LC Analysis

Similar to the treatment of sampling weights in standard log-linear models, we can introduce sampling weights in log-linear LC models. The formula for the log-linear LC model differs from the formula of the standard log-linear model described in equation (1) in that it requires an additional index for the unobserved latent variable(s). Denoting a category of the (joint) latent variable by k, the log-linear LC model with sampling weights can be defined as

$$m_{jk} = \exp(\mathbf{x}_{jk}\boldsymbol{\beta})z_j = h_{jk}\, z_j. \tag{2}$$

Here, m_{jk} is an expected cell frequency before weighting, h_{jk} is an expected cell entry in the population, and z_j is again the inverse of the aggregated sampling weight for answer pattern j. The design matrix with elements $\boldsymbol{x}_{jk}$ defines the log-linear terms appearing in the LC model of interest. As shown below, in the standard LC model these are the main effect, the one-variable terms for the latent and the manifest variables, and the two-variable terms for the relationship between the latent and the manifest variables. Note that m_{jk} and h_{jk} are cell entries in the table including the latent variables. The corresponding entries in the observable table (m_j and h_j) can be obtained by collapsing the table over the index k; that is, $m_j = \sum_k m_{jk}$ and $h_j = \sum_k h_{jk}$. Thus, as in the standard log-linear model, the following relationship holds: $m_j = h_j\, z_j$.

Haberman (1979) proposed the following log-linear LC model:

$$m_{jk} = \exp(\mathbf{x}_{jk}\boldsymbol{\beta})z_{jk} = h_{jk}z_{jk},$$

which is slightly more general than the model we need for dealing with sampling weights. It becomes equivalent to the above model if we set $z_{jk} = z_j$. Models of this form can be estimated with the generally available software packages Newton (Haberman 1988), LEM (Vermunt 1997a), and Latent GOLD (Vermunt and Magidson 2005). In the latter program, $\log(z_{jk})$ should be used as an offset in the module for modeling choice data.

As is shown in the appendix, ML estimation of β parameters of the weighted log-linear LC model under Poisson or multinomial sampling involves solving the following set of likelihood equations (see also Haberman 1979):

$$\sum_{jk} \widehat{m}_{jk}x_{jku} = \sum_{jk} \widehat{n}_{jk}x_{jku}. \tag{3}$$

Here $\widehat{n}_{jk} = n_j\,\widehat{\pi}_{k|j}$, u refers to the log-linear parameter concerned, and $\widehat{\pi}_{k|j} = \widehat{m}_{jk}/\sum_k \widehat{m}_{jk}$ is the estimated probability of being in LC k given response pattern j. Haberman (1979, 1988) showed how to solve this problem using Fisher scoring and a (modified) Newton-Raphson method. The latter method is implemented in the NEWTON program (Haberman 1988). The LEM (Vermunt 1997a) and Latent GOLD (Vermunt and Magidson 2005) programs estimate the weighted log-linear LC model by EM and/or the Newton-Raphson method.

An Inefficient EM Algorithm

Suppose we have an LC model for three indicators Y_1, Y_2, and Y_3 and a single latent variable X. A particular category is referred to by the lower-case equivalents of these symbols: y_1, y_2, y_3, and x. We can now write the log-linear LC model with cell weights as follows:

$$m_{y_1y_2y_3x} = h_{y_1y_2y_3x}\, z_{y_1y_2y_3},$$

with

$$h_{y_1y_2y_3x} = \exp\left(\beta + \beta_x^X + \beta_{y_1}^{Y_1} + \beta_{y_2}^{Y_2} + \beta_{y_3}^{Y_3} + \beta_{y_1x}^{Y_1X} + \beta_{y_2x}^{Y_2X} + \beta_{y_3x}^{Y_3X}\right).$$

For this unrestricted log-linear LC model, the likelihood equations to be solved at iteration cycle t are simply

$$\widehat{m}_{y_1x}^{(t)} = \widehat{n}_{y_1x}^{(t)};\ \widehat{m}_{y_2x}^{(t)} = \widehat{n}_{y_3x}^{(t)};\ \widehat{m}_{y_3x}^{(t)} = \widehat{n}_{y_3x}^{(t)}. \tag{4}$$

The terms at the right-hand side—$\widehat{n}_{y_1x}^{(t)}$, $\widehat{n}_{y_2x}^{(t)}$, and $\widehat{n}_{y_3x}^{(t)}$—are the expected sufficients statistics that should be computed in the E step of the EM algorithm, that is, by the appropriate collapsing of $n_{y_1y_2y_3}\,\widehat{\pi}_{x|y_1y_2y_3}^{(t-1)}$, where $n_{y_1y_2y_3}$ is an observed cell count and $\widehat{\pi}_{x|y_1y_2y_3}^{(t-1)} = \widehat{m}_{y_1y_2y_3x}^{(t-1)} / \sum_x \widehat{m}_{y_1y_2y_3x}^{(t-1)}$ is the posterior probability of belonging to LC x for someone with the corresponding observed values on the three indicators. This quantity is estimated with the parameter values—or with estimated expected cell frequencies—from the previous iteration cycle. In the M step, we need to update the estimates for $\widehat{m}_{y_1y_2y_3x}^{(t)}$, for example, using a set of simple iterative proportional fitting (IPF) cycles. The only modification of an EM algorithm for log-linear LC models without cell weights is that in the current situation the starting values for the expected cell entries—based on the starting values for the parameters—should be multiplied by the cell weights. In other words, the cell weights play a role only before the EM iterations begin.

The main disadvantage of the estimation procedure for the log-linear LC model is that it cannot be used for problems with more than a few indicators since all cells in the estimated cross-tabulation of the observed and latent variables are processed at each iteration step. Hence, a question of interest is whether there is a way to solve the estimation problem by processing only the cells with nonzero observed counts. As in an unweighted analysis, this might be straightforward using the Lazarsfeld and Henry (1968) and Goodman (1974a, 1974b) probability formulation of the LC model instead of the log-linear LC model.

The Probability LC Model With Weights

An unrestricted LC model for three indicators can alternatively be defined using the probability formulation; that is,

$$m_{y_1y_2y_3x} = h_{y_1y_2y_3x}\, z_{y_1y_2y_3},$$

with

$$h_{y_1y_2y_3x} = \gamma\, \pi_x^X \pi_{y_1x}^{Y_1|X}\, \pi_{y_2x}^{Y_2|X}\, \pi_{y_3x}^{Y_3|X}, \tag{5}$$

where π_x^X denotes the unconditional probability of belonging to LC x and $\pi_{y_1x}^{Y_1|X}$ the conditional probability of giving response y_1 on Y_1 given that one belongs to LC x. The terms corresponding to the other two indicators have a similar interpretation. Note that the term γ is included to guarantee that the sample size is reproduced. In an unweighted analysis, γ will be equal to the sample size N. Here, it equals

$$\gamma = \frac{N}{\sum_{y_1 y_2 y_3 x} \pi_x^X \pi_{y_1 x}^{Y_1|X} \pi_{y_2 x}^{Y_2|X} \pi_{y_3 x}^{Y_3|X} z_{y_1 y_2 y_3}}.$$

To distinguish it from the log-linear LC model, we will refer to the model described in equation (5) as the probability LC model.

The probabilities appearing in the probability LC model can be written as a function of the collapsed version of $h_{y_1 y_2 y_3 x}$. For example,

$$\pi_{y_1 x}^{Y_1|X} = \frac{h_{y_1 x}}{\sum_{y_1} h_{y_1 x}}, \tag{7}$$

where $h_{y_1 x} = \sum_{y_2 y_3} h_{y_1 y_2 y_3 x}$, which is proportional to $\exp(\beta_{y_1}^{Y_1} + \beta_{y_1 x}^{Y_1 X})$. This shows the connection between the log-linear and probability LC model. In an unweighted analysis, these probabilities can be estimated using the information from only the nonzero observed cells, which makes it possible to estimate an LC model with many items.[1] A similar result would apply if the likelihood equations to be solved at iteration t could be written as

$$\widehat{m}_{y_1 x}^{(t)} = \sum_{y_2 y_3} \frac{\widehat{n}_{y_1 y_2 y_3 x}^{(t)}}{z_{y_1 y_2 y_3}}; \; \widehat{m}_{y_2 x}^{(t)} = \sum_{y_1 y_3} \frac{\widehat{n}_{y_1 y_2 y_3 x}^{(t)}}{z_{y_1 y_2 y_3}}; \; \widehat{m}_{y_3 x}^{(t)} = \sum_{y_1 y_2} \frac{\widehat{n}_{y_1 y_2 y_3 x}^{(t)}}{z_{y_1 y_2 y_3}}. \tag{8}$$

It can easily be verified that this formula is not equivalent to the conditions defining the ML solution described in equation (4). Actually, working with conditions (8) is equivalent to performing a weighted analysis by using $n_{y_1 y_2 y_3}^{(w)} = \frac{n_{y_1 y_2 y_3}}{z_{y_1 y_2 y_3}}$ as observed frequencies in an unweighted analysis, that is, to using the pseudo-ML estimation approach.

An Efficient EM Algorithm

Now we describe an efficient EM algorithm for the probability LC model with cell weights. For simplicity of exposition, we focus on the estimation of a single set of the model probabilities, namely, $\pi_{y_1 x}^{Y_1|X}$ or, more precisely, $h_{y_1 x}^{Y_1 X}$. Note that because of the relationship described in equation (7), we may redefine the problem of estimating $\pi_{y_1 x}^{Y_1|X}$ as the estimation of $h_{y_1 x}^{Y_1 X}$.

As in the log-linear LC model, in the E step for iteration t, we obtain estimates for the observed marginal frequencies $\widehat{n}_{y_1x}^{(t)}$ (the expected sufficients statistics) using the data and the parameters of the previous iteration. These are then used to obtain new $\widehat{m}_{y_1x}^{(t)}$ in the M step; that is, $\widehat{m}_{y_1x}^{(t)} = \widehat{n}_{y_1x}$. This is all we need to do in an unweighted analysis. In a weighted analysis, however, computation of new $\widehat{m}_{y_1x}^{(t)}$ is not sufficient since the model probabilities are defined in terms of h_{y_1x} rather than in terms of m_{y_1x}. Updated estimates for h_{y_1x} are obtained as follows:

$$\widehat{h}_{y_1x}^{(t)} = \frac{\widehat{m}_{y_1x}^{(t)}}{\widehat{z}_{y_1x}^{(t)}} = \frac{\widehat{n}_{y_1x}^{(t)}}{\widehat{z}_{y_1x}^{(t)}}. \tag{9}$$

As can be seen, computation of the new provisional estimates $\widehat{h}_{y_1x}^{(t)}$ requires provisional estimates for the aggregated cell weights in marginal table $Y_1 - X$, which are denoted by $\widehat{z}_{y_1x}^{(t)}$. It turns out that these can be calculated in the E step together with the $\widehat{n}_{y_1x}^{(t)}$. For that purpose, we first compute the estimated frequencies in the marginal table concerned—$\widehat{m}_{y_1x}^{(t-1)*}$—by collapsing the estimated unweighted frequencies from the previous iteration—$\widehat{m}_{y_1y_2y_3x}^{(t-1)}$—over other variables:

$$\widehat{m}_{y_1x}^{(t-1)*} = \sum_{y_2y_3} \widehat{m}_{y_1y_2y_3x}^{(t-1)} = \sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \; z_{y_1y_2y_3}. \tag{10}$$

At iteration t, the estimates for the cell weights in the marginal table $Y_1 - X$ are

$$\widehat{z}_{y_1x}^{(t)} = \frac{\widehat{m}_{y_1x}^{(t-1)*}}{\widehat{h}_{y_1x}^{(t-1)}}. \tag{11}$$

Thus, a provisional marginal cell weight is the ratio of the current estimates of the unweighted and the weighted marginal cell frequencies. A similar EM algorithm was used by Vermunt (1997b) for log-linear event history models with (partially) unobserved covariates.

Note that $\widehat{m}_{y_1x}^{(t-1)}$ is the provisional value of a cell in the unweighted marginal table concerned based on $\widehat{n}_{y_1x}^{(t-1)}$, whereas $\widehat{m}_{y_1x}^{(t-1)*}$ is the provisional value based on the parameter estimates $\widehat{h}_{y_1x}^{(t-1)}$ and the cell weights. As long as convergence is not reached, $\widehat{m}_{y_1x}^{(t-1)*}$ will not be equal to $\widehat{m}_{y_1x}^{(t-1)}$.

As can be seen from equation (10), for the computation of $\widehat{m}_{y_1x}^{(t-1)*}$ we have to process all cells of the frequency table, which limits the applicability of this approach to LC models with a relatively small number of indicators. An important question is whether it is possible to compute $\widehat{m}_{y_1x}^{(t-1)*}$, and as a consequence, $\widehat{z}_{y_1x}^{(t)}$ using only the nonzero observed cells.

The assumption we make is that the weight for each cell with a zero observed frequency has the same value α; that is, $z_{y_1y_2y_3} = \alpha$ if $n_{y_1y_2y_3} = 0$. Let indicator variable $\varepsilon_{y_1y_2y_3}$ be equal to 1 if $n_{y_1y_2y_3} \neq 0$, and 0 otherwise. Now we can reformulate equation (10) as follows:

$$\widehat{m}_{y_1x}^{(t-1)*} = \sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \, z_{y_1y_2y_3}\varepsilon_{y_1y_2y_3} + \sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \, \alpha\,(1 - \varepsilon_{y_1y_2y_3});$$

That is, we split the formula for $\widehat{m}_{y_1x}^{(t-1)*}$ into two parts, one part corresponding to the nonzero observed cells and the other to the zero observed cells. Making use of the fact that $\sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} = \widehat{h}_{y_1x}^{(t-1)}$, we can rewrite this formula as follows:

$$\begin{aligned} \widehat{m}_{y_1x}^{(t-1)*} &= \sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \, z_{y_1y_2y_3}\varepsilon_{y_1y_2y_3} + \left(\widehat{h}_{y_1x}^{(t-1)}\alpha - \sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \, \alpha \, \varepsilon_{y_1y_2y_3} \right) \\ &= \sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \, (z_{y_1y_2y_3} - \alpha)\varepsilon_{y_1y_2y_3} + \widehat{h}_{y_1x}^{(t-1)}\alpha. \end{aligned}$$

The term $\sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \, (z_{y_1y_2y_3} - \alpha)\varepsilon_{y_1y_2y_3}$ is computed from the nonzero cells, and the term $\widehat{h}_{y_1x}^{(t-1)}$ is a model parameter that is available from the previous iteration cycle; this shows that updating $\widehat{m}_{y_1x}^{(t-1)*}$ and $\widehat{z}_{y_1x}^{(t)}$ requires processing only the nonzero cells.

A question that remains is what value to use for α. Our solution is to use $\alpha = 1$, which amounts to assuming that the weighted and unweighted observed cell frequencies are equal to one another. This is better than $\alpha = 0$ since that gives a model in which the observed zeroes are treated as structural zeroes.[2] With $\alpha = 1$, we get

$$\widehat{m}_{y_1x}^{(t-1)*} = \sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \, (z_{y_1y_2y_3} - 1) + \widehat{h}_{y_1x}^{(t-1)}.$$

Using (11), this implies that $\widehat{z}_{y_1x}^{(t)}$ can be calculated in the E step as follows:

$$\widehat{z}_{y_1x}^{(t)} = \frac{\sum_{y_2y_3} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \left(z_{y_1y_2y_3} - 1\right)}{\widehat{h}_{y_1x}^{(t-1)}} + 1$$

This defines an efficient and quite simple EM algorithm. In the tth E step, we compute new $\widehat{n}_{y_1x}^{(t)}$ and $\widehat{z}_{y_1x}^{(t)}$ using the observed data (frequencies and weights) and the parameters from the previous iteration. In the M step, we use these quantities to obtain new parameter estimates, that is, new $\widehat{h}_{y_1x}^{(t)}$ (see equation [9]).

The γ Parameter

A similar procedure as above has to be used to obtain the γ term, which guarantees that $\sum_{y_1y_2y_3x} \widehat{m}_{y_1y_2y_3x} = N$. Using the definition in equation (6), the estimate for γ at iteration t is simply

$$\widehat{\gamma}^{(t)} = \frac{N}{\sum_{y_1y_2y_3x} \widehat{\pi}_x^{X(t)} \widehat{\pi}_{y_1x}^{Y_1|X(t)} \widehat{\pi}_{y_2x}^{Y_2|X(t)} \widehat{\pi}_{y_3x}^{Y_3|X(t)} z_{y_1y_2y_3}}.$$

Again, it is important to solve this without going through the complete table. It turns out that we can use the following updating scheme for γ:

$$\widehat{\gamma}^{(t)} = \widehat{\gamma}^{(t-1)} \frac{N}{\widehat{m}^{(t-1)*}}.$$

Above, we showed how to obtain $\widehat{m}_{y_1x}^{(t-1)*}$ using only the information from the nonzero cells. In a similar way, we can obtain $\widehat{m}^{(t-1)*}$; that is, by

$$\widehat{m}^{(t-1)*} = \sum_{y_1y_2y_3x} \widehat{h}_{y_1y_2y_3x}^{(t-1)} \left(z_{y_1y_2y_3} - 1\right) + \widehat{h}^{(t-1)}.$$

This computation of γ is relevant not only in an efficient EM algorithm but also for a Newton-Raphson algorithm since the γ term is required to get the correct log-likelihood value.

A Small Simulation Study

How Much Difference Does Weighting Make?

In this section, we show in which types of situations weighting makes sense in an LC analysis, as well as which weighting methods yield asymptotic correct solutions. For the moment, we will work with population distributions in order to prevent sampling fluctuations from influencing the conclusions. More specifically, we assess which weighting methods reproduce the known population values in the case of nonproportional stratified sampling. Later, we investigate how well the methods that work asymptotically (with sample sizes of infinity) work with sample sizes typical for survey research.

The stratification variable is a dichotomous variable denoted by A. The population and sample proportions for $A = 1$—denoted by π_1^A and p_1^A—equal .9 and .5, respectively. This is a typical example of oversampling of a minority group (large firms, large cities, ethnic minorities, females in masculine jobs). We can correct for the oversampling by using sampling weights $w_1^A = 1.8$ and $w_2^A = .2$, respectively.

The population model is an LC model for five dichotomous indicators:

$$\pi_{y_1 y_2 y_3 y_4 y_5 xa}^{Y_1 Y_2 Y_3 Y_4 Y_5 X|A} = \pi_{xa}^{X|A} \pi_{y_1 xa}^{Y_1|XA} \pi_{y_2 xa}^{Y_2|XA} \pi_{y_3 xa}^{Y_3|XA} \pi_{y_4 xa}^{Y_4|XA} \pi_{y_5 xa}^{Y_5|XA},$$

where

$$\pi_{y_r xa}^{Y_r|XA} = \frac{\exp\left(\beta_{y_r a}^{Y_r|A} + \beta_{y_r xa}^{Y_r X|A}\right)}{\sum_{s=1}^{2} \exp\left(\beta_{sa}^{Y_r|A} + \beta_{sxa}^{Y_r X|A}\right)}.$$

We assume that $\pi_{11}^{X|A} = .1$ and $\pi_{12}^{X|A} = .5$, which means that the minority group ($A = 2$) has a much larger probability of belonging to the "low" class ($X = 1$) than the majority group. The overall probability of being in LC 1 is $\pi_1^X = \pi_1^A \pi_{11}^{X|A} + \pi_2^A \pi_{12}^{X|A} = .14$. Note that with $p_1^A = .5$, the unweighted estimate for π_1^X will be .3.

The question is under which conditions are we able to obtain the correct population value of π_1^X. The investigated conditions varied with respect to the specification of the effects-coded log-linear parameters $\beta_{1a}^{Y_r|A}$ and $\beta_{1xa}^{Y_r X|A}$, which define the response probabilities $\pi_{y_r xa}^{Y_r|XA}$. In the *homogeneous* measurement model specification, we set $\beta_{1a}^{Y_r|A} = -.8 + .4\,r$ and $\beta_{11a}^{Y_r X|A} = .5$ for each r (item) and a (group). The *heterogeneous* specifications were obtained by allowing for between-group variation in item difficulty (probability of giving the $Y_r = 2$ response) and/or item discrimination (strength of relationship between X and Y_r): We made one or two items more difficult for $A = 2$ by setting $\beta_{12}^{Y_r|A} = \beta_{11}^{Y_r|A} + .4$ and/or one or two items less discriminating for $A = 2$ by setting $\beta_{112}^{Y_r X|A} = .1$. The eight specifications we used are

I. homogeneous measurement model, $\pi_{y_r xa}^{Y_r|XA} = \pi_{y_r x}^{Y_r|X}$ for all r;
II. Y_3 is more difficult for $A = 2$;
III. Y_3 discriminates less for $A = 2$;
IV. Y_3 is more difficult and discriminates less for $A = 2$;

V. Y_3 is more difficult and Y_4 discriminates less for $A=2$;
VI. Y_3 and Y_4 are more difficult for $A=2$;
VII. Y_3 and Y_4 discriminate less for $A=2$; and
VIII. Y_3 and Y_4 are more difficult and discriminate less for $A=2$.

These are typical measurement models in a multiple-group LC analysis (Clogg and Goodman 1984, 1985).

We use four types of methods to deal with the nonproportional stratified sampling:

0. unweighted analysis;
1. two-step approach consisting of an unweighted analysis with an adjustment of the estimate for π_1^X using the weighted frequencies. This can be either a single-step adjustment of π_1^X (1a) or an iterative reestimation of π_1^X (1b);
2. analysis using weighted frequencies $n_j^{(w)}$;
3. weighted analysis using cell weights z_j; and
4. unweighted analysis with A as a grouping variable assuming either a homogeneous (4a) or a heterogeneous measurement model (4b).

The single-step adjustment after an unweighted analysis (1a) involves reestimating $\widehat{\pi}_x^X$ as follows:

$$\widehat{\pi}_x^X = \sum_j \widehat{\pi}_{k|j} n_j^{(w)}.$$

A problem is that $\widehat{\pi}_{k|j}$ is computed with the wrong unweighted estimate of π_x^X. A better approach seems to be to reestimate π_x^X iteratively, fixing the other model parameters at their estimated values. Reestimation involves using $n_j^{(w)}$ as observed frequencies.

The model that is estimated with methods 0, 1a, 1b, 2, and 3 is a standard LC model with five dichotomous indicators.[3] This means that we do not take into account possible differences in the measurement model across levels of the stratifier. That is the reason that we use the same, possibly incorrect, measurement model in method 4a.

Table 2 presents the results. With a homogeneous measurement model (case I), each of the methods gives the correct value for π_1^X except method 1a. Methods 2 and 3 also give the correct answer if the heterogeneity of the measurement model concerns a single item (cases II–IV), but with heterogeneity in two items the estimate is no longer correct. In the latter cases, the bias of method 3 is slightly smaller than that of method 2. Method 1a works badly in all cases, whereas method 1b works quite well,

Table 2
Estimates for π_1^X Under Eight True Models and Seven Weighting Methods for Dealing With the Stratified Sampling Design

Method	I	II	III	IV	V	VI	VII	VIII
0. Unweighted	.300	.300	.300	.300	.311	.310	.307	.353
1a. Two-step (noniterative)	.218	.202	.226	.212	.222	.198	.236	.244
1b. Two-step (iterative)	.140	.119	.137	.119	.124	.112	.134	.121
2. Pseudo-ML	.140	.140	.140	.140	.144	.142	.141	.151
3. Weighted ML	.140	.140	.140	.140	.143	.141	.141	.148
4a. Multiple group (homogeneous)	.140	.138	.166	.202	.186	.146	.232	.233
4b. Multiple group (heterogeneous)	.140	.140	.140	.140	.140	.140	.140	.140

Note: ML = maximum likelihood.

especially if the group differences concern only item discrimination (cases II and VII). Surprisingly, the inclusion of the stratifier as a grouping variable in combination with the wrong measurement model is among the worst choices, which shows that one has to be cautious with this strategy. The bias of method 4a is, however, small if the heterogeneity concerns only difficulty parameters (cases II and VI). As seen from the results for method 4b, specification of the correct measurement model yields, as might be expected, the correct value for π_1^X in all situations.

From the above, we can conclude that when the measurement model is homogeneous, there is no need to use sampling weights when estimating the class-specific response probabilities. Unbiased estimates for the class sizes (here π_1^X) can be obtained in a second stage in which the weighted frequencies are used as data and the class-specific response probabilities are fixed to the estimated values from the unweighted analysis (method 1b). If the homogeneous measurement model assumption is incorrect for no more than a single item, both methods that use sampling weights (methods 2 and 3) still give the correct estimates for the class sizes. However, in situations where the parameters for more than a single item differ across stratifier levels, these weighting methods will no longer provide unbiased estimates of the class sizes.

Behavior of the Various Weighting Methods With Sampling Fluctuation

Above we studied the asymptotic performance (unbiasedness) of the various weighting methods. In this section, we investigate how well the

Table 3
Means and Standard Deviations for π_1^X Across 1,000 Samples of Size 1,000 Simulated From Eight True Models and Seven Weighting Methods for Dealing With the Stratified Sampling Design

Method	I	II	III	IV	V	VI	VII	VIII
0. Unweighted	.305	.304	.305	.303	.319	.315	.319	.357
	(.046)	(.043)	(.061)	(.052)	(.060)	(.044)	(.079)	(.067)
1a. Two-step	.224	.206	.231	.215	.229	.203	.249	.250
(noniterative)	(.044)	(.039)	(.057)	(.048)	(.055)	(.039)	(.076)	(.064)
1b. Two-step	.150	.128	.149	.128	.137	.121	.155	.136
(iterative)	(.037)	(.032)	(.048)	(.038)	(.044)	(.031)	(.061)	(.049)
2. Pseudo-ML	.160	.156	.162	.155	.167	.158	.165	.172
	(.061)	(.056)	(.069)	(.058)	(.068)	(.056)	(.075)	(.069)
3. Weighted ML	.152	.149	.153	.148	.157	.148	.155	.156
	(.053)	(.049)	(.058)	(.049)	(.056)	(.045)	(.063)	(.057)
4a. Multiple group	.145	.143	.172	.203	.191	.151	.238	.237
(homogeneous)	(.030)	(.028)	(.041)	(.043)	(.040)	(.028)	(.053)	(.037)
4b. Multiple group	.149	.147	.150	.147	.151	.148	.151	.150
(heterogeneous)	(.035)	(.036)	(.040)	(.038)	(.040)	(.036)	(.041)	(.044)

Note: ML = maximum likelihood.

various methods behave when applied to a sample, that is, not only whether estimates are unbiased but also whether they are stable. For this purpose, we use a similar design as in the previous section. The starting points are the same eight populations that differ with respect to the type of heterogeneity of the measurement model across levels of the stratification variable, as well as the various weighting methods described above. The difference is that now we generate samples from the population distributions rather than use population data. The sample size we set to 1,000, a moderate sample size in survey research. The simulation study consists of 1,000 replications.

To summarize, we generate 1,000 samples of size 1,000 from eight populations. To each of these samples, we apply the various methods for dealing with sampling weights. Tables 3 and 4 report the results from this small simulation study; Table 3 gives the average and the standard deviation of the estimated π_1^X value across replications, and Table 4 gives the same information on the logit scale.

The results are similar to the asymptotic results. Overall, the ML weighting method proposed in this article (method 3) performs best, but the difference compared to method 2 (pseudo-ML) is not very large. It can also be

Table 4
Means and Standard Deviations for logit(π_1^X) Across 1,000 Samples of Size 1,000 Simulated Under Eight True Models and Seven Weighting Methods for Dealing With the Stratified Sampling Design

Method	I	II	III	IV	V	VI	VII	VIII
0. Unweighted	−0.832	−0.838	−0.841	−0.845	−0.773	−0.783	−0.786	−0.602
	(0.216)	(0.204)	(0.291)	(0.248)	(0.280)	(0.204)	(0.373)	(0.296)
1a. Two-step	−1.261	−1.368	−1.228	−1.316	−1.239	−1.383	−1.148	−1.127
(noniterative)	(0.250)	(0.243)	(0.323)	(0.283)	(0.315)	(0.245)	(0.414)	(0.343)
1b. Two-step	−1.766	−1.954	−1.789	−1.961	−1.887	−2.012	−1.773	−1.912
(iterative)	(0.294)	(0.294)	(0.376)	(0.342)	(0.373)	(0.289)	(0.471)	(0.417)
2. Pseudo-ML	−1.727	−1.751	−1.718	−1.762	−1.681	−1.729	−1.714	−1.646
	(0.442)	(0.420)	(0.486)	(0.429)	(0.475)	(0.404)	(0.529)	(0.472)
3. Weighted ML	−1.775	−1.796	−1.780	−1.801	−1.743	−1.795	−1.770	−1.746
	(0.403)	(0.380)	(0.439)	(0.388)	(0.427)	(0.349)	(0.475)	(0.420)
4a. Multiple group	−1.792	−1.807	−1.597	−1.392	−1.466	−1.742	−1.188	−1.181
(homogeneous)	(0.241)	(0.231)	(0.288)	(0.273)	(0.265)	(0.218)	(0.304)	(0.207)
4b. Multiple group	−1.768	−1.788	−1.766	−1.786	−1.764	−1.783	−1.761	−1.775
(heterogeneous)	(0.278)	(0.287)	(0.312)	(0.300)	(0.313)	(0.287)	(0.317)	(0.337)

Note: ML = maximum likelihood.

observed that weighting methods 2 and 3 yield larger standard deviations across replications than the other methods, which indicates that there is a clear trade-off between bias and stability of estimates.[4] The fact that weighting increases uncertainty about parameter estimates is a phenomenon reported by various authors (Hendrikx 2002; Winship and Radbill 1994).

Empirical Examples

A Standard LC Model for Six Dichotomous Indicators

To illustrate the performance of various weighting methods with real-life data, we took six dichotomous indicators measuring work values from Dutch samples of the European Values Study (EVS) 1990 and EVS 1999 surveys. The task for respondents was to pick a number of items that they found important in a job out of a list of 15. For this example, we used 6 of the 15 items: Y_1, “chances for promotions”; Y_2, “use initiative”; Y_3, “achieve something”; Y_4, “responsible job”; Y_5, “job interesting”; and Y_6, “meeting abilities.” The Dutch EVS surveys contain case weights in

Table 5
Test Results for the Models Estimated With the EVS Job Attitude Data

		n_j		$n_j^{(w)}$		n_j and z_j	
Model	*df*	L^2	BIC	L^2	BIC	L^2	BIC
One-class	120	2,130.24	1,216.94	2,165.12	1,251.82	2,245.86	1,332.56
Two-class	112	399.50	−452.91	442.73	−409.68	447.09	−405.33
Three-class	104	195.99	−595.54	226.69	−564.84	224.26	−567.27
Four-class	96	121.70	−608.94	154.73	−575.91	157.46	−573.18
Five-class	88	110.17	−559.58	133.62	−536.13	136.15	−533.60

Note: EVS = European Values Study; L^2 = likelihood-ratio statistic; BIC = Bayesian information criterion.

order to correct for the sampling design and for unit nonresponse. Weights differ across age groups, gender, and regions.[5]

The model we assume for the six work-value indicators is an LC model with homogeneous response probabilities across the two time points.[6] Table 5 reports the value of the likelihood-ratio statistic (L^2) and the BIC for one- to five-class models estimated using unweighted frequencies, weighted frequencies, and cell weights. As can be seen, for each of the three methods, the four-class model is the preferred model based on the BIC criterion.[7]

The parameters for the four-class model obtained using the ML approach with cell weights (see Table 6) show that there is one class that finds all items important (class 1) and one class that finds all unimportant (class 4). Classes 2 and 3 take an intermediate position, where class 2 gives higher importance to items Y_2, Y_4, and Y_5 (the self-development items), and the very small class 3 gives higher importance to items Y_1, Y_2, Y_3, and Y_4 (the achievement items).

A similar four-class pattern as in Table 6 is found with both the unweighted analysis and the analysis of weighted frequencies. However, the estimated latent distribution and its change between 1990 and 1999 are somewhat different for the three weighting methods (see Table 7). As can be seen, compared to the unweighted analysis, using the sampling weights increases the size of class 2 and decreases the size of class 3. This effect is stronger when using the ML or cell weights approach described in this article than when using the pseudo-ML or weighted frequencies approach. It can also be seen that the parameter changes, resulting from using the sampling weights, are somewhat larger for the first than for the second time point.

Table 6
Parameters Estimates of the Four-Class Model for the EVS Job Attitude Data Using Cell Weights

	$X=1$	$X=2$	$X=3$	$X=4$
π_x^X	.25	.42	.03	.31
$\pi_{x1}^{X\|T}$	.31	.41	.00	.27
$\pi_{x2}^{X\|T}$	.19	.42	.05	.34
$\pi_{1x}^{Y_1\|X}$	.84	.20	.79	.11
$\pi_{1x}^{Y_2\|X}$	.96	.77	.76	.19
$\pi_{1x}^{Y_3\|X}$	.95	.31	.77	.12
$\pi_{1x}^{Y_4\|X}$	.84	.41	.64	.14
$\pi_{1x}^{Y_5\|X}$	.93	.70	.10	.21
$\pi_{1x}^{Y_6\|X}$	.99	.88	.30	.30

Note: EVS = European Values Study.

Table 7
Estimated Latent Class Proportions of the Four-Class Model for the EVS Job Attitude Data

	Using n_j			Using $n_j^{(w)}$			Using n_j and z_j		
	π_x^X	$\pi_{x1}^{X\|T}$	$\pi_{x2}^{X\|T}$	π_x^X	$\pi_{x1}^{X\|T}$	$\pi_{x2}^{X\|T}$	π_x^X	$\pi_{x1}^{X\|T}$	$\pi_{x2}^{X\|T}$
$X=1$	.25	.30	.20	.25	.31	.19	.25	.31	.19
$X=2$	.37	.35	.39	.40	.40	.40	.42	.41	.42
$X=3$	.09	.08	.11	.06	.03	.09	.03	.00	.05
$X=4$	.29	.27	.30	.29	.26	.32	.31	.27	.34

Note: EVS = European Values Study.

A Nonstandard LC Model

To illustrate the fact that the proposed ML weighting procedure can also be applied with more advanced LC models, we present an example of an LC model for incomplete ranking data. The data are again taken from the Dutch samples of EVS 1990 and EVS 1999. The two indicators of interest form Ingelhart's (post)materialism scale. Respondents select their first and second choices out of the following four "aims of the country": (1) "maintain order in the nation," (2) "more say in important government decisions," (3) "fighting rising prices," and (4) "protect freedom of speech."

We denote the first and second choice by Y_1 and Y_2, respectively. A special feature of a ranking task is that it is impossible to select the same answer twice, which means that the cell counts for $Y_1 = Y_2$ are structurally zero. The data are modeled by a mixture variant of the strict-utility or Bradley-Terry-Luce ranking model (Croon 1989). Using T (time) as a grouping variable, this LC model can be defined as follows:

$$\begin{aligned}\pi_{y_1 y_2 xt}^{Y_1 Y_2 X|T} &= \pi_{xt}^{X|T} \pi_{y_1 xt}^{Y_1|XT} \pi_{y_2 y_1 xt}^{Y_2|Y_1 XT} \\ &= \pi_{xt}^{X|T} \frac{\exp(\beta_{y_1 xt}^{YXT})}{\sum_{y_1} \exp(\beta_{y_1 xt}^{YXT})} \frac{\exp(\beta_{y_2 xt}^{YXT})}{\sum_{y_2 \neq y_1} \exp(\beta_{y_2 xt}^{YXT})},\end{aligned}$$

for $Y_1 \neq Y_2$, and $\pi_{y_1 y_2 xt}^{Y_1 Y_2 X|T} = 0$ otherwise. As can be seen, the first- and second-choice probabilities are parameterized by a set of class- and time-specific utilities β, which are assumed to be equal across choices. For identification, the β parameters are assumed to sum to 0 across alternatives ($\sum_y \beta_{yxt}^{YXT} = 0$).[8] The more complicated subscript $y_2 \neq y_1$ appearing in the sum of the denominator for the second choice is needed because the alternative selected as first choice should be eliminated from the set of alternatives for the second choice.

This mixture Bradley-Terry-Luce ranking model cannot be defined as a log-linear model for the joint distribution of the latent and manifest variables, which means that in this case we have to use the probability formulation of the LC model. An additional feature of the data set we use for this example is that for some respondents the information on the first or second choice is missing. We use the partially observed data in the model estimation, assuming that the missing data is missing at random.

Table 8 reports the test results for the estimated models using the weighting method proposed in this study. The model with the lowest BIC value is the two-class model with a partially heterogeneous measurement structure. Inspection of the estimated parameters of the heterogeneous two-class model shows that the first item, "maintain order in the nation," became more popular between 1990 and 1999, irrespective of the LC. We modeled this by augmenting the homogeneous two-class model by a single parameter capturing the change in popularity of the first item.

The parameter estimates are presented in Table 9. LC 1 is the materialistic class having higher probabilities for Items 1 and 3, and LC 2 is the postmaterialistic class having higher probabilities for Items 2 and 4. The latent change shows an increase of materialistic class, even after filtering out the increased popularity of materialistic Item 1.

Table 8
Test Results for the Models Estimated With the EVS Inglehart Items

Model	L^2	*df*	BIC
One-class heterogeneous	173.18	16	51.40
Two-class heterogeneous	24.57	8	−36.32
One-class homogeneous	250.49	19	105.88
Two-class homogeneous	51.54	14	−55.01
Two-class partially heterogeneous	37.46	13	−61.48

Note: EVS = European Values Study; L^2 = likelihood-ratio statistic; BIC = Bayesian information criterion.

Table 9
Parameter Estimates for Two-Class Partially Heterogeneous LC Model for the EVS Inglehart Items

	$X = 1$	$X = 2$
$\pi_{x1}^{X\|T}$	0.62	0.38
$\pi_{x2}^{X\|T}$	0.66	0.34
β_{1x}^{YX}	1.05	−0.41
β_{2x}^{YX}	−1.06	0.37
β_{3x}^{YX}	−0.33	−1.07
β_{4x}^{YX}	0.33	1.11
β_{12}^{YT}	0.89	0.89
$\pi_{1x1}^{Y_1\|XT}$	0.54	0.12
$\pi_{2x1}^{Y_1\|XT}$	0.07	0.26
$\pi_{3x1}^{Y_1\|XT}$	0.13	0.06
$\pi_{4x1}^{Y_1\|XT}$	0.26	0.55

Note: LC = latent class; EVS = European Values Study.

Discussion

We showed how to obtain ML estimates for log-linear LC models when there are sampling weights by generalizing results from standard log-linear analysis with sampling weights. The method can be implemented using the procedures for log-linear analysis of incomplete tables of Haberman's (1988) NEWTON program and Vermunt's (1997a) LEM program. It can also be implemented in Latent GOLD (Vermunt and Magidson 2005) by using the log of the cell weights as an offset in the module for choice data.

We also showed how to estimate a probability-based LC model with sampling weights without processing the complete table. This makes it possible to apply the proposed weighting method to large problems. The only assumption that needs to be made is that the cell weights are equal for all zero cells, for example, equal to 1.

The probability LC model is a special case of a much broader class of models for frequency tables. Therefore, the proposed weighting method can be generalized quite easily to more complicated probability models with latent and partially observed categorical variables, such as the modified LISREL approach of Hagenaars (1990) and Vermunt (1997b) and the missing data models of Winship and Mare (1989). The proposed method can also be used to define LC models (with many indicators) in which some of the cells are restricted to be structurally zero, such as LC models for capture-recapture data (see Agresti 2002:544).

Because sampling weights increase standard errors, for linear regression analysis Winship and Radbill (1994) recommended using the unweighted solution when parameter estimates are substantively similar with and without weighting. Similar advice could be given for LC analysis: If the variables used to construct the weights do not affect the measurement part of the model, an unweighted analysis is the preferred approach. As we demonstrated, LC sizes can easily be corrected using the two-step approach. In other cases, it is recommended to use sampling weights, where the pseudo-ML and ML weighting methods can be expected to give similar results in terms of parameter values. Whereas in pseudo-ML estimation one can obtain correct standard errors, for instance, by means of a linearization variance estimator or a jackknife procedure (Skinner et al. 1989), construction of valid goodness-of-fit tests is not possible. The main advantage of the proposed ML approach is therefore that it provides valid goodness-of-fit tests.

Appendix
Likelihood Equation for Log-Linear LC Model

In this appendix, we derive the likelihood equation for the weighted log-linear LC model under Poisson and multinomial sampling. The contribution of cell j to Poisson log-likelihood is

$$\log L_j = n_j \log\left(\sum_k m_{jk}\right) - \sum_k m_{jk},$$

and the first-order derivatives are

$$\begin{aligned}\frac{\partial \log L_j}{\partial \beta_u} &= n_j \sum_k \frac{m_{jk}}{\sum_{k'} m_{jk'}} x_{jku} - \sum_k m_{jk} x_{jku} \\ &= \sum_k \widehat{n}_{jk} x_{jku} - \sum_k m_{jk} x_{jku}.\end{aligned}$$

where

$$\widehat{n}_{jk} = \widehat{n}_{jk} \frac{m_{jk}}{\sum_{k'} m_{jk'}} = \widehat{n}_{jk} \pi_{j|k}.$$

This means that the likelihood equations have the form

$$\sum_{jk} m_{jk} x_{jku} = \sum_{jk} \widehat{n}_{jk} x_{jku}.$$

Under multinomial sampling, the contribution of cell j to the likelihood equals

$$\log L_j = n_j \log \left(\sum_k m_{jk} \right) - n_j \log \left(\sum_{j'k} m_{j'k} \right)$$

and the corresponding first-order derivatives have the form

$$\begin{aligned}\frac{\partial \log L_j}{\partial \beta_u} &= n_j \sum_k \frac{m_{jk}}{\sum_{k'} m_{jk'}} x_{jku} - n_j \sum_{j''k} \frac{m_{j''k}}{\sum_{j'k'} m_{j'k'}} x_{j''ku} \\ &= \sum_k \widehat{n}_{jk} x_{jku} - n_j \sum_{j''k} \frac{m_{j''k}}{\sum_{j'k'} m_{j'k'}} x_{j''ku}.\end{aligned}$$

Assuming that $\sum_{jk} m_{jk} = \sum_j n_j$, the sum of all cells (index j) yields

$$\sum_j \frac{\partial \log L_j}{\partial \beta_u} = \sum_{jk} \widehat{n}_{jk} x_{jku} - \sum_{jk} m_{jk} x_{jku}.$$

This shows that Poisson and multinomial sampling give the same likelihood equation.

Note that z_j cancels in the computation of the posterior probabilities:

$$\pi_{k|j} = \frac{m_{jk}}{\sum_{k'} m_{jk'}} = \frac{h_{jk}}{\sum_{k'} h_{jk'}},$$

which shows that the posterior class membership probabilities do not depend on the sampling weights.

Notes

1. It should be noted that latent class (LC) models for large numbers of indicators raise other problems associated with sparseness, which complicates model evaluation.

2. As indicated by one of the reviewers, in some situations one may wish to combine several values of α. More specifically, one may use $\alpha = 0$ for zero observed cells that are considered structural zeroes and $\alpha = 1$ for zero cells that are considered sampling zeroes. This can be easily achieved without any modification of the algorithm described here. Structural zero cells should be given a cell weight of 0 and a $\varepsilon_{y_1 y_2 y_3}$ value of 1 (instead of 0).

3. The model is estimated treating the expected cell counts defining the population as data, that is, as if they were observed frequencies.

4. The mean squared error could be used to determine the combined effect of bias and variability. It can easily be obtained from the numbers reported in these two tables, that is, as the sum of squared bias and squared standard deviation.

5. As was mentioned earlier, the presented maximum-likelihood (ML) approach ignores clustering. But since the European Values Study (EVS) sample for the Netherlands was not clustered, this is valid in the current example. This may, however, not be the case for other EVS countries since each participating country has its own sampling plan.

6. We assume time-homogeneous response probabilities for simplicity of exposition. In the next, more advanced example, we consider models with different response probabilities across time points.

7. It should be noted that the use of L^2 and Bayesian information criterion is somewhat questionable for the pseudo-ML estimation approach (the method based on weighted frequencies) because these measures contain the log-likelihood function in their formulae instead of the pseudo-log-likelihood function.

8. Note that these log-linear parameters are not assumed to sum to zero across LCs. An equivalent model would be obtained by adding this constraint and including an item intercept β_{yt}^{YT} in the model.

References

Agresti, Alan. 2002. *Categorical Data Analysis.* Hoboken, NJ: John Wiley.

Clogg, Cliff C. and Scott R. Eliason. 1987. "Some Common Problems in Log-Linear Analysis." *Sociological Methods & Research* 16:8-44.

Clogg, Cliff C. and Leo A. Goodman. 1984. "Latent Structure Analysis of a Set of Multidimensional Contingency Tables." *Journal of the American Statistical Association* 79: 762-71.

———. 1985. "Simultaneous Latent Structure Analysis in Several Groups." *Sociological Methodology* 15:81-110.

Croon, Marcel A. 1989. "Latent Class Models for the Analysis of Rankings." Pp. 99-121 in *New Developments in Psychological Choice Modeling*, edited by G. De Soete, H. Feger, and K. C. Klauer. Amsterdam: Elsevier Science.

Goodman, Leo A. 1974a. "The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable: Part I—A Modified Latent Structure Approach." *American Journal of Sociology* 79:1179-259.

———. 1974b. "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika* 61:215-31.

Haberman, Shelby J. 1978. *Analysis of Qualitative Data*. Vol. 1, *Introduction Topics.* New York: Academic Press.

———. 1979. *Analysis of Qualitative Data*. Vol. 2, *New Developments.* New York: Academic Press.

———. 1988. "A Stabilized Newton-Raphson Algorithm for Log-Linear Models for Frequency Tables Derived by Indirect Observations." *Sociological Methodology* 18: 193-211.

Hagenaars, Jacques A. 1990. *Categorical Longitudinal Data—Log-linear Analysis of Panel, Trend, and Cohort Data.* Newbury Park, CA: Sage.

Hendrikx, John. 2002. "The Impact of Weights on Standard Errors." Presented at the annual meeting of the Association for Survey Computing, April 17, Imperial College, London, UK.

Lazarsfeld Paul F. and Neil W. Henry. 1968. *Latent Structure Analysis.* Boston: Houghton Mifflin.

Magidson, Jay. 1987. "Weighted Log-Linear Modeling." *Statistical Association 1987 Proceedings of the Social Statistics Section*: 171-4.

Patterson, Blosson H., C. Mitchell Dayton, and Barry I. Graubard. 2002. "Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data."*Journal of the American Statistical Association* 97:721-8.

Rao, J. N. K. and D. Roland Thomas. 1988. "The Analysis of Cross-Classified Categorical Data From Complex Sample Surveys". *Sociological Methodology* 18:213-69.

Skinner, Chris J., Tim Holt, and T. M. Fred Smith, eds. 1989. *Analysis of Complex Surveys*. New York: John Wiley.

Vermunt, Jeroen K. 1997a. *LEM: A General Program for the Analysis of Categorical Data.* Tilburg, Netherlands: Department of Methodology and Statistics, Tilburg University.

———. 1997b. *Log-Linear Models for Event Histories.* Thousand Oaks, CA: Sage.

———. 2002. "Comments on Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data." *Journal of the American Statistical Association* 97:736-7.

Vermunt, Jeroen K. and Jay Magidson. 2005. *Technical Guide to Latent GOLD 4.0: Basic and Advanced.* Belmont, MA: Statistical Innovations.

Wedel, Michel, Frenkel Ter Hofstede, and Jan-Benedict E. M. Steenkamp. 1998. "Mixture Model Analysis of Complex Samples." *Journal of Classification* 15:225-44.

Winship, Christopher and Robert D. Mare. 1989. "Loglinear Models With Missing Data: A Latent Class Approach." *Sociological Methodology* 19:331-67.

Winship, Christopher and Larry Radbill. 1994. "Sampling Weights and Regression Analysis." *Sociological Methods & Research* 23:230-57.

Jeroen K. Vermunt is a professor in the Department of Methodology and Statistics at Tilburg University, Netherlands. He holds a PhD in social sciences from Tilburg University. He has published extensively on categorical data techniques, methods for the analysis of longitudinal and event history data, latent class and finite mixture models, and latent trait models.

Jay Magidson is founder and president of Statistical Innovations, Inc., a Boston-based consulting and software firm. He holds a PhD in managerial economics and decision sciences from Northwestern University. He is widely published in various professional journals including the *Journal of Marketing Research* and *Sociological Methodology*. He was awarded a patent for an innovative graphical display.